

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 12:15:08 ; Search time 17.4 seconds  
(without alignments)  
745.463 Million cell updates/sec

Title: US-09-810-506-2  
Perfect score: 1816  
Sequence: 1 MAPEINTKLTVPVHSATGGE.....FIALSEAGALQYVKAPSAA 335

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	243	13.4	332	1	GLYG_MOUSE	Q9r062 mus musculu
2	241	13.3	349	1	GLYG_HUMAN	P46976 homo sapien
3	240	13.2	501	1	GYG2_HUMAN	O15488 homo sapien
4	235.5	13.0	332	1	GLYG_RABIT	P13280 oryctolagus
5	231	12.7	332	1	GLYG_RAT	O08730 rattus norv
6	165	9.1	371	1	YAOC_SCHPO	Q09680 schizosacch
7	129.5	7.1	380	1	GLG2_YEAST	P47011 saccharomyc
8	122	6.7	618	1	GLG1_YEAST	P36143 saccharomyc
9	110	6.1	286	1	GSPA_BACSU	P25148 bacillus su
10	105	5.8	1251	1	YMI1_YEAST	Q04545 saccharomyc
11	103.5	5.7	2157	1	POLG_HRV1B	P12916 human rhino
12	96.5	5.3	338	1	RFAJ_ECOLI	P27129 escherichia
13	94.5	5.2	443	1	COBB_METJA	Q58816 methanococc
14	93.5	5.1	337	1	RFAI_SALTY	P19816 salmonella
15	93.5	5.1	623	1	HCOE_EURCA	P02242 eurytelma c
16	93	5.1	862	1	LOXA_PHAVU	P27480 phaseolus v
17	92	5.1	630	1	TNPE_STAAR	P37375 staphylococ
18	92	5.1	2151	1	RRPL_HANTV	P23456 hantaan vir
19	91.5	5.0	336	1	RFAJ_SALTY	P19817 salmonella
20	91.5	5.0	342	1	SPDE_LYCES	Q9zs45 lycopersico
21	91	5.0	315	1	SPD1_HYONI	O48658 hyoscyamus
22	90	5.0	372	1	GMDS_HUMAN	O60547 homo sapien
23	89.5	4.9	759	1	ARY2_CALVI	P28514 calliphora
24	88	4.8	824	1	TGLK_RAT	P23606 rattus norv
25	87.5	4.8	324	1	CP21_HORVU	P55747 hordeum vul
26	87	4.8	285	1	Y714_STRCO	Q9x8h0 streptomyc
27	87	4.8	766	1	DPP4_HUMAN	P27487 homo sapien
28	86.5	4.8	570	1	NCAP_MOPEI	P19239 mopeia viru
29	86	4.7	314	1	SPDE_NICSY	O48660 nicotiana s
30	86	4.7	317	1	SPD2_DATST	Q96557 datura stra
31	85.5	4.7	450	1	VD10_BPT5	P11107 bacterioph
32	85.5	4.7	630	1	TNPB_STAAR	P06697 staphylococ
33	85.5	4.7	697	1	YHF0_YEAST	P38721 saccharomyc

34	85.5	4.7	867	1	RRPO_BYDVI	P29044 barley yell
35	85.5	4.7	1314	1	TETX_CLOTE	P04958 clostridium
36	85	4.7	817	1	TGLK_HUMAN	P22735 homo sapien
37	85	4.7	1589	1	CC25_YEAST	P04821 saccharomyc
38	84.5	4.7	578	1	YC20_METJA	Q58617 methanococc
39	84.5	4.7	638	1	NOSZ_PSEST	P19573 pseudomonas
40	84.5	4.7	1147	1	KIN2_YEAST	P13186 saccharomyc
41	83.5	4.6	513	1	AVR2_YEAT	P38444 rattus norv
42	83.5	4.6	724	1	RRPO_TNVD	P27209 tobacco nec
43	83.5	4.6	867	1	RRPO_BYDVP	P09505 barley yell
44	83.5	4.6	1003	1	SYG_CHLMU	Q9plc6 chlamydia m
45	83	4.6	412	1	CCA_ECOLI	P06961 escherichia

ALIGNMENTS

RESULT 1						
GLYG_MOUSE	GLYG_MOUSE	STANDARD;	PRT;	332	AA.	
AC	Q9R062;					
DT	16-OCT-2001 (Rel. 40, Created)					
DT	16-OCT-2001 (Rel. 40, Last sequence update)					
DT	16-OCT-2001 (Rel. 40, Last annotation update)					
DE	Glycogenin-1 (EC 2.4.1.186).					
GN	GYG OR GYGL.					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.					
OX	NCBI_TaxID=10090;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=Skeletal muscle;					
RX	MEDLINE=20011289; PubMed=10542328;					
RA	van Maanen M.-H., Fournier P.A., Palmer T.N., Abraham L.J.;					
RT	"Characterization of mouse glycogenin-1 cDNA and promoter region.";					
RL	Biochim. Biophys. Acta 1447:284-290(1999).					
CC	-!- FUNCTION: SELF-GLUCOSYLATES, VIA AN INTER-SUBUNIT MECHANISM, TO FORM AN OLIGOSACCHARIDE PRIMER THAT SERVES AS SUBSTRATE FOR GLYCOGEN SYNTHASE.					
CC	-!- CATALYTIC ACTIVITY: UDP-glucose + glycogenin = UDP + glucosylglycogenin.					
CC	-!- COFACTOR: SELF-GLUCOSYLATION IS DEPENDENT ON THE PRESENCE OF DIVALENT METAL IONS OF WHICH MANGANESE ION IS THE MOST EFFECTIVE.					
CC	-!- PATHWAY: GLYCOGEN BIOSYNTHESIS.					
CC	-!- SUBUNIT: HOMODIMER, TIGHTLY COMPLEXED TO GLYCOGEN SYNTHASE.					
CC	-!- TISSUE SPECIFICITY: SKELETAL MUSCLE, HEART, IN A LESSER EXTENT IN KIDNEY, LUNG AND BRAIN.					
CC	-!- PTM: SELF-GLUCOSYLATED BY THE TRANSFER OF GLUCOSE RESIDUES FROM UDP-GLUCOSE TO ITSELF, FORMING AN ALPHA-1,4-GLYCAN OF AROUND 10 RESIDUES ATTACHED TO TYR-194.					
CC	-!- SIMILARITY: BELONGS TO THE GLYCOGENIN FAMILY.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).					
CC	-----					
CC	EMBL; AF114031; AAD48901.1; -					
DR	MGD; MGI:1351614; Gyg1.					
DR	InterPro; IPR002495; Glycosyl_transf_8.					
DR	Pfam; PF01501; Glyco_transf_8; 1.					
DR	Transferase; Glycogen biosynthesis; Acetylation; Phosphorylation; Glycoprotein.					
KW	INIT_MET 0	0				BY SIMILARITY.
FT	ACT_SITE 85	85				BY SIMILARITY.
FT	MOD_RES 1	1				ACETYLATION (BY SIMILARITY).
FT	MOD_RES 43	43				PHOSPHORYLATION (BY CAPK) (BY SIMILARITY).
FT	CARBOHYD 194	194				O-LINKED (GLC. . .) (BY SIMILARITY).



QY 80 IEPVYPENQTEFAMAYVYVINYSKLRIWFEFVEYKMIYLDGDIQVFDNIDHLDLPNGQF 139  
Db 62 VLDSGDGAHLTLMKRPGLGVTLTKLHCWSLTQYSKCVFMDADTLVLANIDDLFDREE--- 118  
QY 140 YAVMDCFEKTSWSPQYKIGYCCQCPDKVTWPEAKLGPPLFYFNAGMFVYEPNLSTYH 199  
Db 119 -----LSAAPDP-GWPDG-----FNSGVFVYQSVETYN 146  
QY 200 NLETVKIVPPTLFAEQDFLNMVFK-----DIYKPIPPVYNLVLAMLRHPENIEL--DQ 252  
Db 147 QLLHLASEQGSFDDGQILNFTFSSWATDIRKHLFPFIYNLSSISYSLPAPKVFEGAS 206  
QY 253 KVHVHYCAAGAKPWRFT-----GEEENMDREDIKMLVKKKWDIYNDESIDYKNV 301  
Db 207 AKVHVHFLGR-VKPNWNTYDPRKTSVKSEAHDPNMTHTPEFLIL---WNIFTTNVL----- 257  
QY 302 VIGDSHKKQQTLOQF 316  
Db 258 -----PLLQOF 263  
RESULT 3  
GYG2\_HUMAN STANDARD; PRT; 501 AA.  
ID GYG2\_HUMAN 015489; 015490; 015486; 015485; 015487;  
AC 015488; 015489; 015490; 015486; 015485; 015487;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE GYG2\_HUMAN 2 (EC 2.4.1.186) (GN-2) (GN2).  
GN GYG2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Placentalia; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RC TISSUE=Liver;  
RX MEDLINE=98010589; PubMed=9346895;  
RA Mu J., Skurat A.V., Roach P.J.;  
RT "Glycogenin-2, a novel self-glucosylating protein involved in liver  
RT glycogen biosynthesis.";  
RL J. Biol. Chem. 272:27589-27597(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20184741; PubMed=10721716;  
RA Zhai L., Mu J., Zong H., DePaoli-Roach A.A., Roach P.J.;  
RT "Structure and chromosomal localization of the human glycogenin-2 gene  
RT GYG2.";  
RL Gene 242:229-235(2000).  
RN [3]  
RP CHARACTERIZATION, AND CARBOHYDRATE-LINKAGE SITE TYR-228.  
RX MEDLINE=99074257; PubMed=9857012;  
RA Mu J., Roach P.J.;  
RT "Characterization of human glycogenin-2, a self-glucosylating  
RT initiator of liver glycogen metabolism.";  
RL J. Biol. Chem. 273:34850-34856(1998).  
CC -!- FUNCTION: SELF-GLUCOSYLATES, VIA AN INTER-SUBUNIT MECHANISM, TO  
CC FORM AN OLIGOSACCHARIDE PRIMER THAT SERVES AS SUBSTRATE FOR  
CC GLYCOGEN SYNTHASE.  
CC -!- CATALYTIC ACTIVITY: UDP-glucose + glycogenin = UDP +  
CC glucosylglycogenin.  
CC -!- COFACTOR: SELF-GLUCOSYLATION IS DEPENDENT ON THE PRESENCE OF  
CC DIVALENT METAL IONS OF WHICH MANGANESE ION IS THE MOST EFFECTIVE.  
CC -!- PATHWAY: GLYCOGEN BIOSYNTHESIS.  
CC -!- SUBUNIT: HOMODIMER, TIGHTLY COMPLEXED TO GLYCOGEN SYNTHASE.  
CC -!- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; ALPHA (SHOWN HERE),  
CC BETA, GAMMA, DELTA, EPSILON AND ZETA; ARE PRODUCED BY ALTERNATIVE  
CC SPLICING.  
CC -!- TISSUE SPECIFICITY: EXPRESSED PREFERENTIALLY IN LIVER, HEART, AND  
CC PANCREAS.  
CC -!- PTM: SELF-GLYCOSYLATED BY THE TRANSFER OF GLUCOSE RESIDUES FROM  
CC UDP-GLUCOSE TO ITSELF, FORMING AN ALPHA-1,4-GLYCAN OF AROUND 10

CC RESIDUES ATTACHED TO TYR-228.  
CC -!- SIMILARITY: BELONGS TO THE GLYCOGENIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U94362; AAB84377.1; -.  
DR EMBL; U94363; AAB84378.1; -.  
DR EMBL; U94364; AAB84379.1; -.  
DR EMBL; U94357; AAB84373.1; -.  
DR EMBL; U94358; AAB84374.1; -.  
DR EMBL; U94360; AAB84375.1; -.  
DR EMBL; U94361; AAB84376.1; -.  
DR EMBL; AF179624; AAF61855.1; -.  
DR EMBL; AF179615; AAF61855.1; JOINED.  
DR EMBL; AF179616; AAF61855.1; JOINED.  
DR EMBL; AF179617; AAF61855.1; JOINED.  
DR EMBL; AF179618; AAF61855.1; JOINED.  
DR EMBL; AF179619; AAF61855.1; JOINED.  
DR EMBL; AF179620; AAF61855.1; JOINED.  
DR EMBL; AF179621; AAF61855.1; JOINED.  
DR EMBL; AF179622; AAF61855.1; JOINED.  
DR EMBL; AF179623; AAF61855.1; JOINED.  
DR MIM; 300198; -.  
DR InterPro; IPR002495; Glycosyl\_transf\_8.  
DR Pfam; PF01501; Glyco\_transf\_8; 1.  
KW Transferase; Glycogen biosynthesis; Alternative splicing;  
KW Glycoprotein; Polymorphism.  
FT ACT\_SITE 119 119 BY SIMILARITY.  
FT CARBOHYD 228 228 O-LINKED (GLC..).  
FT VARSPLIC 3 33 MISSING (IN ISOFORM BETA).  
FT VARSPLIC 3 42 MISSING (IN ISOFORM GAMMA).  
FT VARSPLIC 378 448 MISSING (IN ISOFORM DELTA).  
FT VARSPLIC 407 501 MISSING (IN ISOFORM EPSILON).  
FT VARSPLIC 413 448 MISSING (IN ISOFORM ZETA).  
FT VARIANT 270 270 V -> A.  
FT /FTId=VAR\_010401.  
FT MUTAGEN 228 228 Y->F: LOSS OF ACTIVITY.  
FT MUTAGEN 230 230 Y->F: NO LOSS OF ACTIVITY.  
FT CONFLICT 413 413 MISSING (IN REF. 1; AAB84378).  
FT CONFLICT 462 464 EKV -> AGI (IN REF. 1; AAB84376).  
SQ SEQUENCE 501 AA; 55211 MW; 2EDE05FDAD5A7657 CRC64;

Query Match 13.2%; Score 240; DB 1; Length 501;  
Best Local Similarity 25.4%; Pred.No. 2.8e-13;  
Matches 78; Conservative 53; Mismatches 114; Indels 62; Gaps 12;  
QY 22 RAYVTLAGTDYVGVGLAKGLRKAISKYPLVAVLPDVPEDHRKQLVDQGVVKEIE 81  
Db 37 QAFVT-LATNDIYCGALVGLQSLRRHRLTRKLVVLITPQVSSLLR--VILSKVFDEVI 92  
QY 82 PVYPPENQTEFAMAYVYV-----INYSKLRIWFEFVEYKMIYLDGDIQVFDNIDHLDLPN 136  
Db 93 EVNLIDSADYIHLAFLKRPGLGLTLTKLHCWTLTHYKCVFLDADTLVLSNVDELFD--R 150  
QY 137 GQFYAVMDCFCETWSHSPQYKIGYCCQCPDKVTWPEAKLGPPLFYFNAGMFVYEPNLS 196  
Db 151 GEFSAAPD-----PGWPDG-----FNSGVFVQPSLH 177  
QY 197 TYHNLLETVKIVPPTLFAEQDFLNMVFK-----DIYKPIPPVYNLVLAMLRHPENIEL- 250  
Db 178 THKLLQLQHAMEHGSFDDGADQGLLNSFFRNWSTTDIHKHLPFIYNLSSNTMYTSPAFKQF 237  
QY 251 -DQVKVHYCAAGAKPWRFTGEEEN---MDREDIK-----MLVKKWDIYNDESID-YK 299  
Db 238 GSSAKVHVHFLGS-MKPNWNTYDPRKTSVKSEAHDPNMTHTPEFLIL---WNIFTTNVLPLYK 296

QY 300 NVVIGDS 306  
Db 297 SVQAGEA 303

RESULT 4  
GLYG\_RABIT  
ID GLYG\_RABIT STANDARD; PRT; 332 AA.  
AC P13280;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Glycogenin-1 (EC 2.4.1.186).  
GN GYG OR GYG1.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=93100288; PubMed=1281472;  
RA Viskupic E., Cao Y., Zhang W., Cheng C., Depaoli-Roach A.A.,  
RA Roach P.J.;  
RT "Rabbit skeletal muscle glycogenin. Molecular cloning and production  
of fully functional protein in Escherichia coli.";  
RL J. Biol. Chem. 267:25759-25763(1992).  
RN [2]  
RP SEQUENCE.  
RC STRAIN=NEW ZEALAND WHITE; TISSUE=Skeletal muscle;  
RX MEDLINE=90032666; PubMed=2806254;  
RA Campbell D.G., Cohen P.;  
RT "The amino acid sequence of rabbit skeletal muscle glycogenin.";  
RL Eur. J. Biochem. 185:119-125(1989).  
RN [3]  
RP SEQUENCE OF 34-47; 181-201; 209-226 AND 307-324.  
RC TISSUE=Liver;  
RX MEDLINE=89325337; PubMed=2526735;  
RA Smythe C., Villar-Palasi C., Cohen P.;  
RT "Structural and functional studies on rabbit liver glycogenin.";  
RL Eur. J. Biochem. 183:205-209(1989).  
RN [4]  
RP SEQUENCE OF 34-47, AND PHOSPHORYLATION.  
RX MEDLINE=89374676; PubMed=3151442;  
RA Lomako J., Whelan W.J.;  
RT "The occurrence of serine phosphate in glycogenin: a possible  
regulatory site.";  
RL Biofactors 1:261-264(1988).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=94192798; PubMed=8143846;  
RA Alonso M.D., Lomako J., Lomako W.M., Whelan W.J.;  
RT "Tyrosine-194 of glycogenin undergoes autocatalytic glucosylation but  
is not essential for catalytic function and activity.";  
RL FEBS Lett. 342:38-42(1994).  
RN [6]  
RP CHARACTERIZATION.  
RX MEDLINE=95289736; PubMed=771798;  
RA Cao Y., Steinrauf L.K., Roach P.J.;  
RT "Mechanism of glycogenin self-glucosylation.";  
RL Arch. Biochem. Biophys. 319:293-298(1995).  
RN [7]  
RP CHARACTERIZATION, AND MUTAGENESIS OF LYS-85.  
RX MEDLINE=99160387; PubMed=10049511;  
RA Lin A., Mu J., Yang J., Roach P.J.;  
RT "Self-glucosylation of glycogenin, the initiator of glycogen  
biosynthesis, involves an inter-subunit reaction.";  
RL Arch. Biochem. Biophys. 363:163-170(1999).  
CC -!- FUNCTION: SELF-GLUCOSYLATES, VIA AN INTER-SUBUNIT MECHANISM, TO  
CC FORM AN OLIGOSACCHARIDE PRIMER THAT SERVES AS SUBSTRATE FOR  
CC GLYCOGEN SYNTHASE.  
CC -!- CATALYTIC ACTIVITY: UDP-glucose + glycogenin = UDP +  
CC glucosylglycogenin.

CC -!- COFACTOR: SELF-GLUCOSYLATION IS DEPENDENT ON THE PRESENCE OF  
CC DIVALENT METAL IONS OF WHICH MANGANESE ION IS THE MOST EFFECTIVE.  
CC -!- PATHWAY: GLYCOGEN BIOSYNTHESIS.  
CC -!- SUBUNIT: HOMODIMER TIGHTLY COMPLEXED TO THE 86 KDA CATALYTIC  
CC SUBUNIT OF GLYCOGEN SYNTHASE.  
CC -!- PTM: SELF-GLYCOSYLATED BY THE TRANSFER OF GLUCOSE RESIDUES FROM  
CC UDP-GLUCOSE TO ITSELF, FORMING AN ALPHA-1,4-GLYCAN OF AROUND 10  
CC RESIDUES ATTACHED TO TYR-194.  
CC -!- SIMILARITY: BELONGS TO THE GLYCOGENIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L01791; AAA31404.1; -.  
DR PIR; S06482; S06482.  
DR PIR; A45094; A45094.  
DR InterPro; IPR002495; Glycosyl\_transf\_8.  
DR Pfam; PF01501; Glyco\_transf\_8; 1.  
KW Transferase; Glycogen biosynthesis; Acetylation; Phosphorylation;  
KW Glycoprotein.  
FT INIT\_MET 0 0  
FT ACT\_SITE 85 85  
FT MOD\_RES 1 1  
FT MOD\_RES 43 43  
FT CARBOHYD 194 194  
FT MUTAGEN 85 85  
FT CONFLICT 38 38  
FT CONFLICT 88 88  
FT CONFLICT 97 97  
SQ SEQUENCE 332 AA; 37266 MW; DD64F6DE6CC5A8FA CRC64;  
  
Query Match 13.0%; Score 235.5; DB 1; Length 332;  
Best Local Similarity 23.8%; Pred. No. 4.1e-13;  
Matches 85; Conservative 49; Mismatches 138; Indels 85; Gaps 15;  
  
QY 22 RAYVTFLAGTDYVKGWVGLAKGLRKAKSKYPLWAVLPDVPEDHRK--QLVDQGCVVKE 79  
Db 3 QAFVT-LTTNDAYAKGALVGLSSLKQHRTRRLAVLTTPQVSDTRKALEIVFDEIVTD 61  
QY 80 IEPVYPPEQTEFAMAYVYVINYSKLRIWEFVEYNKMIYLDGDIQVFDNIDHFLDLPNGQF 139  
Db 62 ILDSGDSAHLTLMKRPELGVTLTKLHCWSLTQYSKCVFMDADTLVLANIDDLFEREE--- 118  
QY 140 YAVMDCFEKTNWSHSPQYKIGYCCQCPDKVTWPEAKLGPPLFYNAGMFVYEPNLSTYH 199  
Db 119 -----LSAAPDP-GWPDG-----FNSGVFVYQPSVETYN 146  
QY 200 NLLETVKIVPPTLFAEQDFLNMVFK-----DIYKPIPPVYNLVLAHLWRHPENIEL--DQ 252  
Db 147 QLLHVASEQGSFDDGQGLLNTFFNSWATTDIRKHLPIFYNLSSISYSLPAFAKAFGAN 206  
QY 253 VKVWHYCAAGAKPWRFT-----GEENMDREDIKMLVKKWDIYN----- 292  
Db 207 AKVWHFLGQ-TKPNWYTYDTTKTSVRSEGHDPMTHT--PQFLNVWWDIFTTSVPLLOQ 262  
QY 293 ----DESLDYKNV--VIGD-SH---KKQOTLQQFIEALS-----EAGALQYVKAPS 333  
Db 263 FGLVQDTCVQHVHEDVSGAVSHLSLGETPATTPFFVSSEERKERWEQQAQADYMGADS 319  
  
RESULT 5  
GLYG\_RAT  
ID GLYG\_RAT STANDARD; PRT; 332 AA.  
AC O08730;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)







KW Complete proteome.  
FT CONFLICT 12 12 C -> A (IN REF. 4).  
FT CONFLICT 27 27 S -> V (IN REF. 4).  
SQ SEQUENCE 286 AA; 33522 MW; A282BE6E3B6518DE CRC64;  
  
Query Match 6.1%; Score 110; DB 1; Length 286;  
Best Local Similarity 23.5%; Pred.No. 0.027;  
Matches 62; Conservative 39; Mismatches 101; Indels 62; Gaps 13;  
  
QY 62 VPEDHRKQLVDQ---GCVVKEIEPVYPPEQTQEFAM-----AYVINYSKLRIWEF 109  
Db 47 IKPDNKKRLEETLTKFGVPIEFLE---VDTNMYEHAVERSHITKAAAYRISIPDLIKDES 103  
QY 110 VEYNKMIYLDGDIQVFDNIDHLFDLPNGQF--YAVMDCFCCKTWSHSPQYKIGYCCQCPD 167  
Db 104 IK--RMIYIDCDALVLEDISKLDLDIAPYTVAAVEDA-----GQHERLKE 147  
QY 168 KVTWPEAKLGPKPPPLYFNAGMFVYEPNLSTYHNLTQVIV-----PPTLFA---EQDF 218  
Db 148 MNVDTGK-----YFNSGIMIIDFESWRKQKQNTITE--KVINFINEHPDEDFLVLHDQDA 198  
QY 219 LNMYFKDIYKPIPPVYN---LVL-----AMLWRHPENIELDQVWVHYCAAGAKPWRP 268  
Db 199 LNAILDQWYELHPRWNAQTYIMLKLPSTLLGRKQYNETRENPAIVHFC-GGEKPWNS 257  
QY 269 TGEENMDREDIKMLVKKWDIYN 292  
Db 258 NTKHPYRDEYFHYMSYTKWNTIGN 281  
  
RESULT 10  
YML\_YEAST  
ID YML\_YEAST STANDARD; PRT; 1251 AA.  
AC Q04545; Q03621;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 141.5 kDa zinc finger protein in TUB1-CPR3 intergenic region.  
GN YML081W.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE OF 1-488 FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Gentles S., Bowman S., Barrell B.G., Rajandream M.A.;  
RL Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE OF 469-1251 FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Brown D., Bowman S., Barrell B.G., Rajandream M.A.;  
RL Submitted (OCT-1994) to the EMBL/GenBank/DDBJ databases.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- SIMILARITY: STRONG, TO YEAST ZMS1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z46660; CAA86657.1; -.  
DR EMBL; Z46373; CAA86497.1; -.  
DR HSSP; P07248; 1ARD.  
DR SGD; S0004546; YML081W.  
DR InterPro; IPR000822; Znf-C2H2.  
DR Pfam; PF00096; zf-C2H2; 2.  
DR PRINTS; PR00048; ZINCINGER.

RESULT 9  
GSPA\_BACSU  
ID GSPA\_BACSU STANDARD; PRT; 286 AA.  
AC P25148;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE General stress protein A.  
GN GSPA OR IPA-12D.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=92216127; PubMed=1806041;  
RA Glaser P., Kunst F., Debarbouille M., Vertes A., Danchin A., Dedonder R.;  
RA "A gene encoding a tyrosine tRNA synthetase is located near sacs in Bacillus subtilis."  
RT DNA Seq. 1:251-261(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=95020537; PubMed=7934828;  
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W., Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I., Presecan E., Santana M., Schneider E., Schweizer J., Vertes A., Rapoport G., Danchin A.;  
RA "Bacillus subtilis genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees."  
RT Mol. Microbiol. 10:371-384(1993).  
RN [3]  
RP SEQUENCE OF 110-286 FROM N.A.  
RC STRAIN=QB1072;  
RX MEDLINE=90373338; PubMed=2116367;  
RA Zukowski M.M., Miller L., Cosgwell P., Chen K., Aymerich S., Steinmetz M.;  
RA "Nucleotide sequence of the sacs locus of Bacillus subtilis reveals the presence of two regulatory genes."  
RT Gene 90:153-155(1990).  
RN [4]  
RP CHARACTERIZATION, AND SEQUENCE OF 1-28.  
RX MEDLINE=95286551; PubMed=7768864;  
RA Antelmann H., Bernhardt J., Schmid R., Hecker M.;  
RA "A gene at 333 degrees on the Bacillus subtilis chromosome encodes the newly identified sigma B-dependent general stress protein Gspa."  
RT J. Bacteriol. 177:3540-3545(1995).  
RL -!- INDUCTION: BY DIFFERENT STRESSES SUCH AS HEAT SHOCK AND SALT STRESS AND BY STARVATION.  
CC -!- SIMILARITY: TO H.INFLUENZAE HI0259.  
CC -!- SIMILARITY: TO LIPOPOLYSACCHARIDE 1,3-GALACTOSYLTRANSFERASE (REAI) AND TO LIPOPOLYSACCHARIDE 1,2-GLUCOSYLTRANSFERASE (REAJ).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X52480; CAA36721.1; ALT\_INIT.  
DR EMBL; X73124; CAA51568.1; -.  
DR EMBL; M29333; AAA75337.1; -.  
DR EMBL; Z99123; CAB15869.1; -.  
DR PIR; S16423; S16423.  
DR Subtilist; BGI0558; gspa.  
DR InterPro; IPR002495; Glycosyl\_transf\_8.  
DR Pfam; PF01501; Glyco\_transf\_8; 1.









```
Db 289 ADI-----LYIGGYPELFKEELSRNKENMIESIKFDFGYIGECGGL-MYITKSIDNVPM 342
QY 233 VYNL-VLAWLRHPENIELDQKVHVHYCAAGAKPWRTGEE-----ENMDREDIKMLVK 285
Db 343 VGLNCSAVMTKHKVQGLSVYKAEFLDCLIGRKGLKFKGHEFHYSKLVNIKEERFAYKIE 402
QY 286 KWDIYND-ESLDYKNVIGDSHKKQQTQLQQFIEALSEAG 324
Db 403 RGRGIINLDGIFNGKVLAGYLHNHAVANPYFASSMVNFG 442

RESULT 14
RFAI_SALTY
ID RFAI_SALTY STANDARD; PRT; 337 AA.
AC P19816; O68267;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide 1,3-galactosyltransferase (EC 2.4.1.44).
GN RFAI OR WAAI OR STM3718.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RX MEDLINE=98204873; PubMed=9535865;
RA Heinrichs D.E., Monteiro M.A., Perry M.B., Whitfield C.;
RT "The assembly system for the lipopolysaccharide R2 core-type of
RT Escherichia coli is a hybrid of those found in Escherichia coli K-12
RT and Salmonella enterica. Structure and function of the R2 Waak and
RT Waal homologs.";
RL J. Biol. Chem. 273:8849-8859(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP PRELIMINARY SEQUENCE OF 1-197 FROM N.A.
RA Flock J.-I.;
RL Submitted (JUL-1990) to the EMBL/GenBank/DBJ databases.
RN [4]
RP PRELIMINARY SEQUENCE OF 198-337 FROM N.A.
RC STRAIN-LT2;
RX MEDLINE=91045080; PubMed=2235496;
RA Carstenius P., Flock J.-I., Lindberg A.;
RT "Nucleotide sequence of rfaI and rfaJ genes encoding
RT lipopolysaccharide glycosyl transferases from Salmonella
RT typhimurium.";
RL Nucleic Acids Res. 18:6128-6128(1990).
CC -1- FUNCTION: ADDS THE GALACTOSE(I) GROUP ON THE GLUCOSE(I) GROUP
CC OF LPS.
CC -1- CATALYTIC ACTIVITY: UDP-galactose + lipopolysaccharide = UDP + D-
CC galactosyl-lipopolysaccharide.
CC -1- PATHWAY: LIPOPOLYSACCHARIDE CORE BIOSYNTHESIS.
CC -1- SIMILARITY: TO E.COLI RFAI AND TO RFAJ.
CC -1- CAUTION: REF.3 AND REF.4 SEQUENCES DIFFER FROM THAT SHOWN DUE TO
CC FRAMESHIFTS AND OTHER SEQUENCING ERRORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF026386; AAC16412.1; -.
DR EMBL; AE008873; AAL22577.1; -.
DR EMBL; X53847; CAA37841.1; ALT_FRAME.
DR PIR; S12097; S12097.
DR StyGene; SG10337; rfaI.
DR InterPro; IPR002495; Glycosyl_transf_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
KW Lipopolysaccharide biosynthesis; Glycosyltransferase; Transferase;
KW Complete proteome.
FT CONFLICT 125 125 R -> RE (IN REF. 1).
FT CONFLICT 210 210 E -> Q (IN REF. 1).
SQ SEQUENCE 337 AA; 38905 MW; 6DB1FA1531AE25F7 CRC64;

Query Match 5.1%; Score 93.5; DB 1; Length 337;
Best Local Similarity 20.3%; Pred. No. 0.88;
Matches 47; Conservative 36; Mismatches 83; Indels 65; Gaps 11;

QY 84 YPENQTEF-----AMAYVYVINYSKLRI-----WEFVEY-----NKM 115
Db 67 FDESDQQRFEALAKQYATQIVVYLIDCERLKLSPSTKNWTYATYFRFIADYFSDKTDV 126
QY 116 IYLDGDIQVFDNIDHFLDL--PNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPE 173
Db 127 LYLDADIACKGSIQELIDLNFAENEAIAVW-AEGELEW-----WTK 166
QY 174 AKLGPKPP---LYFNAGMFVYEPNLSTYHN-----LLETVKIVPPTLFAEQDFLNMY 222
Db 167 RSVSLATPGLVSGYFNAGFILINPLWTAENISKKAIEMLKDPVVQRITHLDQDVLNIF 226
QY 223 FKDIYKPIPPVYNLVLAHLWRHPENI--ELDQVKV-VHYCAAGAKPWRTG 270
Db 227 LVNKARFVDKKFNTQFSLNYELKDSVINPVDATVVFVHYIGP-TKPWHSWG 276

RESULT 15
HCYE_EURCA
ID HCYE_EURCA STANDARD; PRT; 623 AA.
AC P02242;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hemocyanin E chain (HcE).
GN HCE.
OS Eurytelma californica (American tarantula).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Mygalomorphae; Theraphosidae; Aphonopelma.
OX NCBI_TaxID=29932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90319102; PubMed=2371273;
RA Voll W., Voit R.;
RT "Characterization of the gene encoding the hemocyanin subunit e from
RT the tarantula Eurytelma californicum.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:5312-5316(1990).
RN [2]
RP SEQUENCE.
RX MEDLINE=84059635; PubMed=6357986;
RA Schneider H.-J., Drexel R., Feldmaier G., Linzen B., Lottspeich F.,
RA Henschen A.;
RT "Hemocyanins in Spiders, XVIII. Complete amino-acid sequence of
RT subunit e from Eurytelma californicum hemocyanin.";
RL Hoppe-Seyler's Z. Physiol. Chem. 364:1357-1381(1983).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91060544; PubMed=2246235;
RA Voit R., Feldmaier-Fuchs G.;
RT "Arthropod hemocyanins. Molecular cloning and sequencing of cDNAs
RT encoding the tarantula hemocyanin subunits a and e.";
```

J. Biol. Chem. 265:19447-19452(1990).  
[4]  
SEQUENCE OF 74-599 FROM N.A.  
MEDLINE=86300721; Pubmed=3017715;  
Voit R., Schneider H.-J.;  
"Tarantula hemocyanin mRNA. In vitro translation, cDNA cloning and  
nucleotide sequence corresponding to subunit e.";  
Eur. J. Biochem. 159:23-29(1986).  
-!- FUNCTION: HEMOCYANINS ARE COPPER-CONTAINING OXYGEN CARRIERS  
OCCURRING FREELY DISSOLVED IN THE HEMOLYMPH OF MANY MOLLUSKS AND  
ARTHROPODS.  
-!- SUBUNIT: TARANTULA HEMOCYANIN IS A 24-CHAIN POLYMER WITH SEVEN  
DIFFERENT CHAINS IDENTIFIED.  
-!- SUBCELLULAR LOCATION: Extracellular.  
-!- TISSUE SPECIFICITY: HEMOLYMPH.  
-!- MISCELLANEOUS: THE TWO COPPER IONS BOUND EACH HAVE 3 NITROGEN  
LIGANDS (PRESUMABLY CONTRIBUTED BY HIS RESIDUES) AND SHARE A  
BRIDGING LIGAND (POSSIBLY CONTRIBUTED BY A TYR RESIDUE) IN  
ADDITION TO BINDING OXYGEN.  
-!- SIMILARITY: BELONGS TO THE TYROSINASE FAMILY. HEMOCYANIN  
SUBFAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X16894; CAA34772.1; -;  
DR EMBL; X04291; CAA27838.1; -;  
DR EMBL; X16650; CAA34643.1; -;  
DR EMBL; X16651; CAA34643.1; JOINED.  
DR EMBL; X16652; CAA34643.1; JOINED.  
DR EMBL; X16653; CAA34643.1; JOINED.  
DR EMBL; X16654; CAA34643.1; JOINED.  
DR EMBL; X16655; CAA34643.1; JOINED.  
DR EMBL; X16656; CAA34643.1; JOINED.  
DR EMBL; X16657; CAA34643.1; JOINED.  
DR PIR; S06701; BHTLE.  
DR HSSP; P04253; IOXY.  
DR InterPro; IPR000896; Hemocyanin.  
DR InterPro; IPR002227; Tyrosinase.  
DR Pfam; PF00372; hemocyanin; 1.  
DR PRINTS; PR00187; HAEMOCYANIN.  
DR PROSITE; PS00209; HEMOCYANIN\_1; 1.  
DR PROSITE; PS00210; HEMOCYANIN\_2; 1.  
DR PROSITE; PS00498; TYROSINASE\_2; 1.  
KW Transport; Oxygen transport; Copper; Glycoprotein; Hemolymph.  
FT INIT\_MET 0  
FT METAL 168 168 COPPER 1 (PROBABLE).  
FT METAL 172 172 COPPER 1 (PROBABLE).  
FT METAL 199 199 COPPER 1 (PROBABLE).  
FT METAL 319 319 COPPER 2 (PROBABLE).  
FT METAL 323 323 COPPER 2 (PROBABLE).  
FT METAL 359 359 COPPER 2 (PROBABLE).  
FT DISULFID 528 576 BY SIMILARITY.  
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CONFLICT 29 29 D -> R (IN REF. 2).  
FT CONFLICT 46 46 C -> D (IN REF. 2).  
FT CONFLICT 79 79 R -> K (IN REF. 2).  
FT CONFLICT 90 90 A -> H (IN REF. 2).  
FT CONFLICT 110 110 MISSING (IN REF. 2).  
FT CONFLICT 159 159 K -> R (IN REF. 1 AND 3).  
FT CONFLICT 216 216 R -> H (IN REF. 2).  
FT CONFLICT 254 254 H -> M (IN REF. 2).  
FT CONFLICT 307 307 F -> H (IN REF. 2).  
FT CONFLICT 325 325 M -> MK (IN REF. 2).  
FT CONFLICT 528 530 CGW -> DGK (IN REF. 2).  
FT CONFLICT 561 563 NGH -> D (IN REF. 2).  
FT CONFLICT 602 602 F -> L (IN REF. 3).  
FT CONFLICT 608 608 V -> P (IN REF. 3).

SQ SEQUENCE 623 AA; 71545 MW; E411052A80814004 CRC64;  
Query Match 5.1%; Score 93.5; DB 1; Length 623;  
Best Local Similarity 26.0%; Pred. No. 1.9;  
Matches 45; Conservative 24; Mismatches 73; Indels 31; Gaps 11;  
QY 123 QVFDNIDHLDLPNGQFYAVMDCFCCEKTSWSPQ-YKIGYCCQCPDKVTWPEAKLGPKPP 181  
| : || || | : : || | : : || | : : || | : : || | : : || :  
Db 27 QIDARLHHLGRLPQGELEFS---CFHEEDLEAEATELYKILYAKDFDEV---INLAKQSR 79  
QY 182 LYFNAGMFVYEPNLSYHNLLLETVK-IVPPTLFAEQDFLNMVFKDIYKPIPPVYNLVLM 240  
| : || || | : : || | : : || | : : || | : : || | : : || :  
Db 80 TFFVNEGLFVYAVSVALLHR--DDCKGIVVPA-----IQEIFFPDRFVPTETI-NLAVKE 129  
QY 241 LWRHPENIELDQVKVHYCAAGAKPWRTGEENMD--REDIKMLVKKW-WDI 290  
| : || || | : : || | : : || | : : || | : : || | : : || :  
Db 130 AANHP-----DQDISVHVETG----NILDEEYKLAYFKEDVGVTNAHHWHHI 173

Search completed: July 2, 2002, 12:15:10  
Job time: 372 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 12:14:44 ; Search time 48.67 Seconds  
(without alignments)  
1190.740 Million cell updates/sec

Title: US-09-810-506-2  
Perfect score: 1816  
Sequence: 1 MAPEINTKLTVPVHSATGGE.....FIEALSEAGALQYVKAPSAA 335

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL19:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1816	100.0	335	10 Q9FXB2	Q9fxb2 arabidopsis
2	1516.5	83.5	334	10 O80518	O80518 arabidopsis
3	1390.5	76.6	344	10 O22893	O22893 arabidopsis
4	1388	76.4	341	10 Q9XEJ7	Q9xej7 brassica na
5	1345.5	74.1	334	10 Q9XGG4	Q9xgn4 pisum sativ
6	1319	72.6	328	10 Q40710	Q40710 oryza sativ
7	1298.5	71.5	334	10 O22693	O22693 arabidopsis
8	1284.5	70.7	333	10 Q9XGN4	Q9xgn4 ajuga rept
9	1219	67.1	221	10 Q945L1	Q945L1 arabidopsis
10	1216.5	67.0	292	10 Q9XGN3	Q9xgn3 ajuga rept
11	1198.5	66.0	361	10 Q9STQ9	Q9stq9 arabidopsis
12	1198	66.0	333	10 Q9FFA1	Q9ffa1 arabidopsis
13	1110	61.1	345	10 O80766	O80766 arabidopsis
14	1053.5	58.0	318	10 Q947G8	Q947g8 lycopersico
15	763.5	42.0	213	10 Q9SPE1	Q9spel vitis ripar
16	258.5	14.2	1201	10 O80649	O80649 arabidopsis

17	248	13.7	307	5	Q9W2J6	Q9w2j6 drosophila
18	248	13.7	655	10	Q940B5	Q940b5 arabidopsis
19	248	13.7	659	10	Q9LSB1	Q9lsb1 arabidopsis
20	232	12.8	557	10	Q9FZ37	Q9fz37 arabidopsis
21	225.5	12.4	372	3	O43061	O43061 schizosacch
22	216	11.9	614	10	Q94HG3	Q94hg3 oryza sativ
23	205.5	11.3	260	2	Q53234	Q53234 rhodobacter
24	205	11.3	333	10	Q9SZB0	Q9szb0 arabidopsis
25	197	10.8	429	5	Q22997	Q22997 caenorhabdi
26	194	10.7	284	5	Q95Q50	Q95q50 caenorhabdi
27	194	10.7	300	5	O45819	O45819 caenorhabdi
28	189	10.4	291	16	Q9L8S6	Q9l8s6 rhizobium m
29	186.5	10.3	283	12	Q9J879	Q9j879 spodoptera
30	181	10.0	276	12	Q99GV2	Q99gv2 heliocoverp
31	178	9.8	276	12	Q9E236	Q9e236 helicoverpa
32	166	9.1	269	12	Q91F08	Q91f08 cydia pomon
33	158.5	8.7	277	12	Q9PZ00	Q9pz00 xestia c-ni
34	157.5	8.7	404	16	Q97P73	Q97p73 streptococc
35	157.5	8.7	546	10	O04031	O04031 arabidopsis
36	147.5	8.1	442	10	O23503	O23503 arabidopsis
37	140	7.7	460	3	Q9Y761	Q9y761 kluyveromyc
38	140	7.7	470	10	Q9ZQP4	Q9zqp4 arabidopsis
39	135.5	7.5	289	12	Q9E7P3	Q9e7p3 spodoptera
40	134	7.4	306	2	Q50948	Q50948 neisseria g
41	134	7.4	307	2	Q9REX5	Q9rex5 neisseria g
42	134	7.4	376	3	O43062	O43062 schizosacch
43	132	7.3	398	16	Q97P77	Q97p77 streptococc
44	131	7.2	406	16	Q97P76	Q97p76 streptococc
45	129.5	7.1	311	2	P96945	P96945 neisseria m

ALIGNMENTS

RESULT 1  
Q9FXB2  
ID Q9FXB2 PRELIMINARY; PRT; 335 AA.  
AC Q9FXB2;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PUTATIVE GALACTINOL SYNTHASE (AT1G56600/F25P12\_16).  
GN F25P12.95.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
RA Altafi H., Nguyen M., Lam B., Southwick A., Bel, Buehler E., Chin C.,  
RA Chiou J., Choi E., Dunn P., Gonzalez A., Howng B., Kim C., Koo T.,  
RA Lee J.M., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,  
RA Pham P., Sakano H., Schwartz J., Shinn P., Thaveri A., Toriumi M.,  
RA Vaysberg M., Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;  
RA Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,  
RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,  
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S., Miranda M., Narusaka M.,  
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,  
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,  
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
RA Ecker J.R.;  
RT "Arabidopsis ORF clones."  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,  
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,

Query Match	100.0%;	Score 1816;	DB 10;	Length 335;
Best Local Similarity	100.0%;	Pred. No. 1.8e-161;		
Matches 335;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	MAPEINTKLTVPVHSATGGEKRAYVTFLAGTGDYVKGVGLAKGLRKA	AKSKYPLVVA	VLP 60	
Db 1	MAPEINTKLTVPVHSATGGEKRAYVTFLAGTGDYVKGVGLAKGLRKA	AKSKYPLVVA	VLP 60	
QY 61	DVPEDHRKQLVDQGCVVKEIEPVYPENQTEFAMAYVINYSKLRIWEFVEYNKMIYLDG	120		
Db 61	DVPEDHRKQLVDQGCVVKEIEPVYPENQTEFAMAYVINYSKLRIWEFVEYNKMIYLDG	120		
QY 121	DIQVFDNIDHLFDLPNGQFYAVMDCFCEKTSWHSPOYKIGYCQCQCPDKVTWP	EAKLGP	KP 180	
Db 121	DIQVFDNIDHLFDLPNGQFYAVMDCFCEKTSWHSPOYKIGYCQCQCPDKVTWP	EAKLGP	KP 180	
QY 181	PLYFNAGMFVYEPNLSTYHNLLLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVLAM	240		
Db 181	PLYFNAGMFVYEPNLSTYHNLLLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVLAM	240		
QY 241	LWRHPENIELDQVKVHYCAAGAKPWRTGEEENMDREDIKMLVKKWWDIYNDES	LDYKN 300		
Db 241	LWRHPENIELDQVKVHYCAAGAKPWRTGEEENMDREDIKMLVKKWWDIYNDES	LDYKN 300		
QY 301	VWIGDSHKKQQTLOQFIEALSEAGALQYVKAPSA	335		
Db 301	VWIGDSHKKQQTLOQFIEALSEAGALQYVKAPSA	335		
RESULT 2				
O80518	PRELIMINARY;	PRT;	334 AA.	
ID O80518				
AC O80518;				
DT 01-NOV-1998 (TrEMBLrel. 08, Created)				
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)				
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)				
DE F14J9.1 PROTEIN (WATER STRESS INDUCED PROTEIN-LIKE PROTEIN).				
GN F14J9.1.				
OS Arabidopsis thaliana (Mouse-ear cress).				
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.				
OX NCBI_TaxID=3702;				
RN [1]				
RP SEQUENCE FROM N.A.				
RC STRAIN=CV. COLUMBIA;				
RA Federspiel N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,				
RA Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,				
RA Oji O., Osborne B.I., Shinn P., Sun H., Toriumi M., Vysotskaia V.S.,				
RA Yu G., Ecker J., Theologis A., Davis R.W.;				
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.				
RN [2]				
RP SEQUENCE FROM N.A.				
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,				
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,				
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,				
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,				



RP SEQUENCE FROM N.A.  
RA Yamada K., Banh J., Banno F., Dale J.M., Goldsmith A.D., Lee J.M.,  
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Yamamura Y., Yu G.,  
RA Yu S., Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C.,  
RA Koesema E., Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M.,  
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,  
RA Southwick A., Tracy S.E., Shinozaki K., Davis R.W., Ecker J.R.,  
RA Theologis A.;  
RT "Full Length cDNA of gene t8i13.2/At2g47180 (GI:2275196).";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC002337; AAB63818.1; -;  
DR EMBL; AY056139; AAL07218.1; -;  
DR InterPro; IPR002495; Glycosyl\_transf\_8.  
DR Pfam; PF01501; Glyco\_transf\_8; 1.  
SQ SEQUENCE 344 AA; 39596 MW; 46784EA16DBD3A46 CRC64;

Query Match 76.6%; Score 1390.5; DB 10; Length 344;  
Best Local Similarity 74.7%; Pred. No. 1.1e-121;  
Matches 257; Conservative 31; Mismatches 47; Indels 9; Gaps 2;

QY 1 MAPEI-----NTKLTVPVHSATGGEKRAYVTFLAGTDYVGVGLAKGLRKAISKY 52  
Db 1 MAPGLTQTADAMSTVITITKPSLPSVQSDRAYVTFLAGNGDYVGVGLAKGLRKAISKY 60

QY 53 PLVAVLPDVPEDHRKQLVDQGCVVKEIEPVYPPEQTQFAMAYVINYSLRIWFEVEY 112  
Db 61 PLVAVLPDVPEDHRRILVDQGCIVREIEPVYPPEQTQFAMAYVINYSLRIWFEVEY 120

QY 113 NKMIYLDGDIQVFDNIDHLDLPNGQFYAVMDCFCCEKTWSHSPQYKIGYCCQCPDKVTWP 172  
Db 121 SKMIYLDGDIQVYENIDHLDLPDGLYAVMDCFCCEKTWSHTPYKIRYCCQCPDKVQWP 180

QY 173 EAKLGPKPLPYFNAGMFVYEPNLSYHNLETVKIVPPTLFAEQDFLNMFKDIYKPIPP 232  
Db 181 KAELEGPPALYFNAGMFLYEPNLETVEDLLRTLKITPTTFAEQDFLNMFKKIYKPIPL 240

QY 233 VYNLVAMLWRHPENIELDQKVVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYN 292  
Db 241 VYNLVAMLWRHPENVELGKVVHYCAAGSKPWRFTGKEANMEREDIKMLVKKWWDIYD 300

QY 293 DESLDYKN-VVIGDSHKKQOTLQOFIEALSEAGALQYVKAPSA 335  
Db 301 DESLDYKKPVTVVDTEVDLVNLPKPFITALTAEAGRLNVTAPSA 344

RESULT 4  
Q9XEJ7 PRELIMINARY; PRT; 341 AA.  
AC Q9XEJ7;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE GALACTINOL SYNTHASE (FRAGMENT).  
GN GS.  
OS Brassica napus (rape).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Brassica.  
OX NCBI\_TaxID=3708;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Georges F., Bock C.A., Hussain A.K., Yan W., Keller W.A.;  
RT "Cloning and Characterization of a Full Length cDNA for Galactinol  
RT Synthase from Brassica napus";  
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF106954; AAD26116.1; -;  
DR InterPro; IPR002495; Glycosyl\_transf\_8.  
DR Pfam; PF01501; Glyco\_transf\_8; 1.  
FT NON\_TER 1  
SQ SEQUENCE 341 AA; 39059 MW; 12ED74936583D157 CRC64;

Query Match 76.4%; Score 1388; DB 10; Length 341;  
Best Local Similarity 76.4%; Pred. No. 1.9e-121;  
Matches 253; Conservative 35; Mismatches 35; Indels 8; Gaps 3;

QY 7 TKLTVPVHSATGGEKRAYVTFLAGTDYVGVGLAKGLRKAISKYPLVAVLPDVPEDH 66  
Db 17 TKPSPPVHG-----DRAYVTFLAGNGDYVGVGLAKGLRKAISKYPLVAVLPDVPEDH 71

QY 67 RKQLVDQGCVVKEIEPVYPPEQTQFAMAYVINYSLRIWFEVEYKNMIYLDGDIQVFD 126  
Db 72 RRVLVEQGCIVREIEPVYPPEQTQFAMAYVINYSLRIWFEVEYKNMIYLDGDIQVYE 131

QY 127 NIDHLDLPNGQFYAVMDCFCCEKTWSHSPQYKIGYCCQCPDKVTWPEAKLGPKPLYPFNA 186  
Db 132 NIDHLDLPDGYFYAVMDCFCCEKTWSHTPYKIGYCCQCPDKVTWPEAKLGPKPLYPFNA 191

QY 187 GMFVYEPNLSYHNLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVAMLWRHPE 246  
Db 192 GMFVFEPLDLYEDLLRTLKITPTTFAEQDFLNMFKDIYKPIPPVYNLVAMLWRHPE 251

QY 247 NIELDQKVVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDESIDYKNVWIGDS 306  
Db 252 NVELDKVVHYCAAGSKPWRFTGKEANMEREDIKMLVKKWWDIYNDDSLDYKKSVDGL 310

QY 307 HKKQQ--TLQOFIEALSEAGALQYVKAPSA 335  
Db 311 VEESDVNLPKPFISALTAEAGPVKYVTAPSA 341

RESULT 5  
Q9XGG4 PRELIMINARY; PRT; 334 AA.  
AC Q9XGG4;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE PUTATIVE GALACTINOL SYNTHASE (EC 2.4.1.123).  
OS Pisum sativum (Garden pea).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.  
OX NCBI\_TaxID=3888;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=EMBRYO (AXIS AND COTYLEDONS);  
RA Jones R.G.; Harrison C.J.; Hedley C.L.;  
RT "Cloning of Galactinol synthase from Pea embryo";  
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ243815; CAB51130.1; -;  
DR InterPro; IPR002495; Glycosyl\_transf\_8.  
DR Pfam; PF01501; Glyco\_transf\_8; 1.  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 334 AA; 38696 MW; BF5522AC35A2D651 CRC64;

Query Match 74.1%; Score 1345.5; DB 10; Length 334;  
Best Local Similarity 72.9%; Pred. No. 1.8e-117;  
Matches 245; Conservative 45; Mismatches 43; Indels 3; Gaps 3;

QY 1 MAPEINTKLTVPVHSATGGEKRAYVTFLAGTDYVGVGLAKGLRKAISKYPLVAVLP 60  
Db 1 MAPEIVQTSTKPVGTGT-KLKRAYVTFLAGNGDYVGVGLAKGLRKAISKYPLVAVLP 59

QY 61 DVPEDHRKQLVDQGCVVKEIEPVYPPEQTQFAMAYVINYSLRIWFEVEYKNMIYLDG 120  
Db 60 DVPEHREMLESQGCIVREIQPVYPPEQTQFAMAYVINYSLRIWFEVEYKNMIYLDG 119









RA Liu S., Li J., Araujo R., Au M., Brendel V., Buehler E., Conway A.B.,  
RA Conway A.R., Dewar K., Feng J., Kim C., Kurtz D., Li Y., Palm C.J.,  
RA Shinn P., Sun H., Davis R.W., Ecker J.R., Federspiel N.A.,  
RA Theologis A.;  
RT "Arabidopsis thaliana chromosome 1 BAC T13D8 sequence.";  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. COLOMBIA;  
RA Theologis A.;  
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC004473; AAC24075.1; -;  
DR InterPro; IPR002495; Glycosyl\_transf\_8.  
DR Pfam; PF01501; Glyco\_transf\_8; 1.  
SQ SEQUENCE 345 AA; 39372 MW; B4EDDE0FDA075E69 CRC64;  
  
Query Match 61.1%; Score 1110; DB 10; Length 345;  
Best Local Similarity 59.9%; Pred. No. 1.9e-95;  
Matches 203; Conservative 46; Mismatches 64; Indels 26; Gaps 3;  
  
QY 20 EKRAYVTFAGTGDYVGVGLAKGLRKAASKYPLVAVLPDVPEDHRKQLVDQGVVKE 79  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
10 KERAYVTFAGNGDYVGVGLAKGLRKVKSAAYPLVAVMLPDVPEEHREILRSQGCIVRE 69  
  
QY 80 IEPVYPPEQTEFAMAYVYINYSKLRIW-----EFVEYNKMIYL 118  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
70 IEPVHPDSDAYARAYIYINYSKLRIWNVSVYIYRLHENESLRLSLNNFEYNKMIYL 129  
  
QY 119 DGDIOQVFNIDHFLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPEAKLGP 178  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
130 DADIOQVFNIDHFLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPEAKLGP 189  
  
QY 179 KPPLYFNAGMFVEYEPNLSYHNLTETVKIVPTLFAEQDFLNMFKDIYKPIPPVYNLV 238  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
190 PPSPYFNAGMFVEYEPNLSYHNLTETVKIVPTLFAEQDFLNMFKDIYKPIPPVYNLV 249  
  
QY 239 AMLWRHPENIELDQVVKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDES 298  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
250 SVLWRHPGKVDLESVKVHYCPGPKRYTGEEENMDREDIKMLVKKWWDIYNDES 309  
  
QY 299 KNVVIGDSHKKQTLQ--FIEALSEAGALQYVKAPSAA 335  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
310 KPKSPAD---LEATVLESTIIASVTEAPLSYSPAAPSAA 345  
  
RESULT 14  
Q947G8 PRELIMINARY; PRT; 318 AA.  
AC Q947G8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE PUTATIVE GALACTINOL SYNTHASE 1 (EC 2.4.1.123).  
GN GOLS-1.  
OS Lycopersicon esculentum (Tomato).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Downie B., Dahal P., Nonogaki H., Gurusinghe S., Yim K., Fukanaga K.,  
RA Alvarado V., Bradford K.J.;  
RT "Galactinol synthase gene expression in tomato seeds.";  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF311943; AAL26804.1; -;  
KW Transferase; Glycosyltransferase.  
SQ SEQUENCE 318 AA; 36402 MW; AE66D71E45097E52 CRC64;

Query Match 58.0%; Score 1053.5; DB 10; Length 318;  
Best Local Similarity 58.4%; Pred. No. 3.3e-90;

Matches 202; Conservative 48; Mismatches 57; Indels 39; Gaps 8;  
  
QY 1 MAPEI--NTKLTVPVHSATGGEKRAYVTFAGTGDYVGVGLAKGLRKAASKYPLVAV 58  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
1 MAPEFESGTMATTIQKSS---CAYVTFAGNGDYVGVGLAKGLRKAASKYPLVAV 56  
  
QY 59 LPDVPEDHRKQLVDQGVVKEIEPVYPP-ENQTEFAMAYVYINYSKLRIWEFVEYNKMIY 117  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
57 LPDVPPEHRMILTRHGCIVKEIEPLAPLSQSLDKYARSYYVLNYSKLRIWEFVEYSKMVY 116  
  
QY 118 LDGDIQVFNIDHFLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPEAKLG 177  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
117 LDGDMQVFNIDHFLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPEAKLG 164  
  
QY 178 PKPPLYFNAGMFVEYEPNLSYHNLTETVKIVPTLFAEQDFLNMFKDIYKPIPPVYNLV 237  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
165 PRPSVYFNAGMFVFPNPSVYVRLNLTLLKVTPTQFAEQDFLNMFKDIYKPIPYTYNML 224  
  
QY 238 LAMLRHPENIELDQVVKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDES 297  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
225 LAMLRHPEKIEVNRKAKAVHYCSPGAKPKYTKGEEHMDREDIKMLVKKWWDIYN 284  
  
QY 298 YKNVIGDSHKKQTLQ--FIEALSEAGA-----LQYVKAPSAA 335  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
285 -----HKAQGST---VEANRLRGAASFDTNISALYITSPSAA 318  
  
RESULT 15  
Q9SPE1 PRELIMINARY; PRT; 213 AA.  
AC Q9SPE1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE GALACTINOL SYNTHASE (FRAGMENT).  
GN WSI76.  
OS Vitis riparia (Frost grape) (Vitis vulpina).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;  
OC Vitis.  
OX NCBI\_TaxID=96939;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=FLOWER BUDS;  
RA Li X.-Z., McKersie B.D.;  
RT "Freezing tolerance in grape flower buds.";  
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF178569; AAD55726.1; -;  
DR InterPro; IPR002495; Glycosyl\_transf\_8.  
DR Pfam; PF01501; Glyco\_transf\_8; 1.  
FT NON\_TER 1  
SQ SEQUENCE 213 AA; 23590 MW; 6D008A67FF729EEA CRC64;  
  
Query Match 42.0%; Score 763.5; DB 10; Length 213;  
Best Local Similarity 80.9%; Pred. No. 2.4e-63;  
Matches 144; Conservative 12; Mismatches 15; Indels 7; Gaps 2;  
  
QY 15 SATGGEK-----RAYVTFAGTGDYVGVGLAKGLRKAASKYPLVAVLPDVPEDHRK 68  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
5 SATGLAKAASISSRAYVTFAGNGDYVGVGLAKGLRKVKVTAYPLVAVLPDVPAEHRR 64  
  
QY 69 QLVDQGVVKEIEPVYPPENQTEFAMAYVYINYSKLRIWEFVEYNKMIYLDGDIQVFN 128  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
65 ILEDQGVVREIEPVNPPDNQTFAMAYVYINYSKLRIWEFVEYSKMIYLDGDIQVFGNI 124  
  
QY 129 DHLFDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPEAKLGPPLFYNA 186  
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :  
125 DHLFDLDDGYFYAVMDCFCCKTWSNSPQYKIGYCCQCPDKVQCP-AEMGPAPPLYFNA 181  
  
Search completed: July 2, 2002, 12:14:44

Job time: 376 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 12:09:33 ; Search time 29.88 Seconds  
(without alignments)  
1077.307 Million cell updates/sec

Title: US-09-810-506-2  
Perfect score: 1816  
Sequence: 1 MAPEINTKLTVPVHSATGGE.....FIEALSEAGALQYVKAPSAA 335

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1816	100.0	335	2 G96607	probable galactino
2	1516.5	83.5	334	2 F86226	hypothetical prote
3	1390.5	76.6	344	2 A84912	probable galactino
4	1319	72.6	328	2 T07610	WSI76 protein - ri
5	1298.5	71.5	334	2 H96629	hypothetical prote
6	1198.5	66.0	361	2 T06009	hypothetical prote
7	1110	61.1	345	2 T02295	hypothetical prote
8	258.5	14.2	1201	2 T00444	hypothetical prote
9	241	13.3	333	1 JC4695	glycogenin glucosy
10	235.5	13.0	332	1 A45094	glycogenin glucosy
11	232	12.8	557	2 H96590	hypothetical prote
12	225.5	12.4	372	2 T40489	hypothetical prote
13	205.5	11.3	260	2 A45729	glycogenin homolog
14	205	11.3	333	2 T05984	hypothetical prote
15	197	10.8	429	2 T31001	hypothetical prote
16	194	10.7	300	2 T25275	hypothetical prote
17	165	9.1	371	2 S55490	hypothetical prote
18	157.5	8.7	404	2 C95206	glycosyl transfera
19	157.5	8.7	546	2 G86221	protein F7G19.14 l
20	147.5	8.1	442	2 A71433	hypothetical prote
21	143.5	7.9	321	2 T24773	hypothetical prote
22	140	7.7	470	2 A84772	probable glycojeni
23	134	7.4	376	2 T40488	hypothetical prote
24	132	7.3	398	2 G95205	glycosyl transfera
25	131	7.2	406	2 H95205	glycosyl transfera
26	129.5	7.1	380	2 S55176	GLG2 protein - yea
27	118	6.5	114	2 JC6077	P13 protein - Leuc
28	110	6.1	286	2 S16423	general stress pro
29	110	6.1	491	2 S58330	probable membrane

30	108.5	6.0	517	2 A71707	lipopolysaccharide
31	108.5	6.0	696	2 D95206	hypothetical prote
32	108	5.9	814	2 A95206	glycosyl transfera
33	107	5.9	335	2 H91191	hypothetical prote
34	107	5.9	335	2 A86039	probable LPS biosy
35	106	5.8	394	2 A71963	probable lipopolys
36	105	5.8	1251	2 S49645	probable membrane
37	103.5	5.7	2157	1 GNNY1B	genome polyprotein
38	101	5.6	572	2 T51525	hypothetical prote
39	97.5	5.4	175	2 T15067	hypothetical prote
40	97.5	5.4	311	2 AE2277	hypothetical prote
41	96.5	5.3	337	2 AH0972	lipopolysaccharide
42	96.5	5.3	338	2 S47847	lipopolysaccharide
43	96.5	5.3	624	1 BHTLE	hemocyanin chain e
44	95	5.2	4589	2 T14914	dynain beta heavy
45	94.5	5.2	443	2 D64477	cobyric acid a,c

ALIGNMENTS

RESULT 1

G96607

probable galactinol synthase F25P12.95 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: G96607

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719

A;Accession: G96607

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-335 <STO>

A;Cross-references: GB:AE005173; NID:g9954752; PIDN:AAG09103.1; GSPDB:GN00141

C;Genetics:

A;Gene: F25P12.95

A;Map position: 1

Query Match 100.0%; Score 1816; DB 2; Length 335;  
Best Local Similarity 100.0%; Pred. No. 4.2e-143;  
Matches 335; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAPEINTKLTVPVHSATGGEKRAYVTFLAGTDGYVGVGLAKGLRKAKSKYPLVAVLP 60

|||||

Db 1 MAPEINTKLTVPVHSATGGEKRAYVTFLAGTDGYVGVGLAKGLRKAKSKYPLVAVLP 60

Qy 61 DVPEDHRKQLVDQGCVVKEIEPVYPPENQTEFAMAYYVINYSKLRIWEFVEYNKMIYLDG 120

|||||

Db 61 DVPEDHRKQLVDQGCVVKEIEPVYPPENQTEFAMAYYVINYSKLRIWEFVEYNKMIYLDG 120

Qy 121 DIQVFDNIDHFLDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCQCQCPDKVTWPEAKLGPKP 180

|||||

Db 121 DIQVFDNIDHFLDLPNGQFYAVMDCFCCKTWSHSPQYKIGYCQCQCPDKVTWPEAKLGPKP 180

Qy 181 PLYFNAGMFVYEPNLSTYHNLETVKIVPPTLFAEQDFLNMVFKDIYKPIPPVYNLVAM 240

|||||

Db 181 PLYFNAGMFVYEPNLSTYHNLETVKIVPPTLFAEQDFLNMVFKDIYKPIPPVYNLVAM 240

Qy 241 LWRHPENIELDQVKVVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDESLDYKN 300

|||||

Db 241 LWRHPENIELDQVKVVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDESLDYKN 300

Qy 301 VVIGDSHKKQOTLQQFIEALSEAGALQYVKAPSAA 335

Db 301 VVIGDSHKKQQTLOQFIEALSEAGALQYVKAPSA 335

RESULT 2

F86226

hypothetical protein [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: F86226

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719

A;Accession: F86226

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-334 <STO>

A;Cross-references: GB:AE005172; NID:g3482910; PIDN:AAC33195.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

Query Match 83.5%; Score 1516.5; DB 2; Length 334;

Best Local Similarity 80.9%; Pred. No. 3.1e-118;

Matches 275; Conservative 30; Mismatches 24; Indels 11; Gaps 2;

QY 1 MAPEINTKLTVPVHSATGGEKRAYVTFLAGTGDYVKGVGLAKGLRKAASKYPLVAVLP 60

Db 1 MAPEMNKLSY-----GEKKRAYVTFLAGTGDYVKGVGLAKGLRKAASKYPLVAVLP 54

QY 61 DVPEDHRKQLVDQGCVVKEIEPVYPPEQTAFAMAYVINYSLRIWEFVEYNKMIYLDG 120

Db 55 DVPADHRRQLLDQGCVIKEIQVYPPDNQTFAMAYVINYSLRIWEFVEYSKLIYLDG 114

QY 121 DIQVFDNIDHFLDPLNGQFYAVMDCFCCEKTSWSPQYKIGYCCQCPDKVTWPEAKLGPKP 180

Db 115 DIQVFENIDHFLDPLDNGFYAVKDCFCCEKTSWSPQYKIGYCCQCPDKVTWPESELGPKP 174

QY 181 PLYFNAGMFVYEPNLSTYHNLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVAM 240

Db 175 PLYFNAGMFVYEPNLSTYHNLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVAM 234

QY 241 LWRHPENIELDQVKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDESIDYKN 300

Db 235 LWRHPENIELNEAKVHYCAAGAKPWRFTGEEENMDREDIKMLVERWWDIYNDESIDYKN 294

QY 301 VVI-----GDSHKKQQTLOQFIEALSEAGALQYVKAPSA 335

Db 295 FNVHCGQKEDVHRKPKTLPOFFTDLSEADVLCQAKAPSA 334

RESULT 3

A84912

probable galactinol synthase [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C;Accession: A84912

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487

A;Accession: A84912

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-344 <STO>

A;Cross-references: GB:AE002093; NID:g2275196; PIDN:AAB63818.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g47180

A;Map position: 2

Query Match 76.6%; Score 1390.5; DB 2; Length 344;

Best Local Similarity 74.7%; Pred. No. 9.3e-108;

Matches 257; Conservative 31; Mismatches 47; Indels 9; Gaps 2;

QY 1 MAPEI-----NTKLTVPVHSATGGEKRAYVTFLAGTGDYVKGVGLAKGLRKAASKY 52

Db 1 MAPGLTQTADAMSTVTITKPSLPSVQDSDRAYVTFLAGNGDYVKGVGLAKGLRKAASKY 60

QY 53 PLVVAVLPDVPEDHRKQLVDQGCVVKEIEPVYPPEQTAFAMAYVINYSLRIWEFVEY 112

Db 61 PLVVAVLPDVPEDHRKQLVDQGCVVKEIEPVYPPEQTAFAMAYVINYSLRIWEFVEY 120

QY 113 NKMIYLDGDIQVFDNIDHFLDPLNGQFYAVMDCFCCEKTSWSPQYKIGYCCQCPDKVTWP 172

Db 121 SKMIYLDGDIQVFDNIDHFLDPLNGQFYAVMDCFCCEKTSWSPQYKIRYCCQCPDKVQWP 180

QY 173 EAKLGPKPPLYFNAGMFVYEPNLSTYHNLETVKIVPPTLFAEQDFLNMFKDIYKPIPP 232

Db 181 KAEELGEPALYFNAGMFVYEPNLSTYHNLETVKIVPPTLFAEQDFLNMFKDIYKPIPP 240

QY 233 VYNLVAMLWRRHPENIELDQVKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYN 292

Db 241 VYNLVAMLWRRHPENIELDQVKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYN 300

QY 293 DESLDYKN-VVIGDSHKKQQTLOQFIEALSEAGALQYVKAPSA 335

Db 301 DESLDYKKPVTVDTEVDLVNLPKPFITALEAGRLNVVTPASAA 344

RESULT 4

T07610

WS176 protein - rice

C;Species: Oryza sativa (rice)

C;Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 21-Jul-2000

C;Accession: T07610

R;Takahashi, R.; Joshee, N.; Kitagawa, Y.

Plant Mol. Biol. 26, 339-352, 1994

A;Title: Induction of chilling resistance by water stress, and cDNA sequence analysis

A;Reference number: S52642; MUID:95036005

A;Accession: T07610

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-328 <TAK>

A;Cross-references: EMBL:D26537; NID:g454879; PIDN:BAA05538.1; PID:g537404

A;Experimental source: strain Somewake

A;Note: induced by water stress

Query Match 72.6%; Score 1319; DB 2; Length 328;

Best Local Similarity 71.0%; Pred. No. 7.5e-102;

Matches 238; Conservative 41; Mismatches 48; Indels 8; Gaps 2;

QY 1 MAPEINTKLTVPVHSATGGEKRAYVTFLAGTGDYVKGVGLAKGLRKAASKYPLVAVLP 60

Db 2 MGPVSVSE---KKALAAARRAYVTFLAGDGDYVKGVGLAKGLRVRSAAYPLVAVLP 57

QY 61 DVPEDHRKQLVDQGCVVKEIEPVYPPEQTAFAMAYVINYSLRIWEFVEYNKMIYLDG 120

Db 58 DVPGEHRRKLVQGCVVREIQVYPPESQTFAMAYVINYSLRIWEFVEYERMVYLD 117

QY 121 DIQVFDNIDHFLDPLNGQFYAVMDCFCCEKTSWSPQYKIGYCCQCPDKVTWPEAKLGPKP 180

Db 118 DIQVFDNIDHFLDPLDNGFYAVKDCFCCEKTSWSPQYKIGYCCQCPDEAVWERELGPPP 177

QY 181 PLYFNAGMFVYEPNLSTYHNLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVAM 240

Db 178 PLYFNAGMFVHEPGLGTAKDLDALVVTPTTFAEQDFLNMFFREQYKPIPNVYNLVLAM 237  
QY 241 LWRHPENIELDQVKVVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDESLDYKN 300  
Db 238 LWRHPENVLDQVKVVHYCAAGSKPWRFTGKEENMDREDIKMLVKRWWDIYNDESLDYKE 297  
QY 301 VVIGDSHKKQOTLQQFIEALSEAGALQYVKAPSAA 335  
Db 298 ----EEDNADEASQPMRTALAEAGAVKYFPAPSAA 328

RESULT 5  
H96629  
hypothetical protein F8A5.2 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C;Accession: H96629  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marzialli,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719  
A;Accession: H96629  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-334 <STO>  
A;Cross-references: GB:AE005173; NID:g2462751; PIDN:AAB71970.1; GSPDB:GN00141  
C;Genetics:  
A;Gene: F8A5.2  
A;Map position: 1

Query Match 71.5%; Score 1298.5; DB 2; Length 334;  
Best Local Similarity 70.7%; Pred. No. 3.9e-100;  
Matches 241; Conservative 35; Mismatches 52; Indels 13; Gaps 4;  
QY 1 MAPEINTK---LTVPVHSATGGCKRAYVTFLAGTGDYVKGVVGLAKGLRKAASKYPLVVA 57  
Db 1 MAPEISVNPMYLSEKAHQAP--PRRAYVTFLAGNGDYVKGVVGLAKGLRKAASKYPLVVA 58  
QY 58 VLPDVPEDHRKQLVDQGVVKEIEPVYPPEQTFFAMAYVINYSKLRIWEFVEYNKMIY 117  
Db 59 MLPDVPPEHREILRSQGVVREIEPVYPPDNQVEFAMAYVINYSKLRIWNEEYSKMIY 118  
QY 118 LDGDIQVFDNIDHFLDLPNGQFYAVMDCFCEKTSWSPQYKIGYCCQCPDKVTWPEAKLG 177  
Db 119 LDADIQVFDNIDHFLDLSDAYFYAVMDCFCEKTSWLSQYISIGYCCQCPDKVTWPEAKLG 178  
QY 178 PKPPLYFNAGMFVYEPNLSTYHNLLLETVKIVPPTLFAEQDFLNMFFKDIYKPIPPVYNLV 237  
Db 179 PPPLYFNAGMFVFEPSPLTYESLLQTLTITPPSPFAEQDFLNMFFKDIYKPIPPVYNLV 238  
QY 238 LAMLWRHPENIELDQVKVVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDESLD 297  
Db 239 LAMLWRHPENVELEKVKVVHYCAAGSKPWRFTGEEANMDREDIKMLVDKWDVYNDESLD 298  
QY 298 YKNVVIGDSHK---KQOTLQQFIEALSEAGALQYVKAPSAA 335  
Db 299 FKSIPADAEETVTKSILASVLEP-----EMTYFPAPSAA 334

RESULT 6  
T06009  
hypothetical protein T25K17.60 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999

C;Accession: T06009  
R;Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, March 1999  
A;Reference number: Z15382  
A;Accession: T06009  
A;Molecule type: DNA  
A;Residues: 1-361 <BEV>  
A;Cross-references: EMBL:AL049171; GSPDB:GN00062; ATSP:T25K17.60  
A;Experimental source: cultivar Columbia; BAC clone T25K17  
C;Genetics:  
A;Gene: ATSP:T25K17.60  
A;Map position: 4  
A;Introns: 136/3; 244/3; 289/3

Query Match 66.0%; Score 1198.5; DB 2; Length 361;  
Best Local Similarity 65.0%; Pred. No. 8.6e-92;  
Matches 221; Conservative 31; Mismatches 61; Indels 27; Gaps 2;  
QY 21 KRAYVTFLACTGDYVKGVVGLAKGLRKAASKYPLVAVLPDVPEDHRKQLVDQGVVKEI 80  
Db 24 KRAYVTFLAGNKDYWMGVVGLAKGLRKAASKYPLVAVLPDVPEDHRQILLAQGCIIREI 83  
QY 81 EPVYPPENQTEFAMAYVINYSKLRIWE-----FVEYNKM 115  
Db 84 EPVYPPENKTYGSMAYVINYSKLRIWEKNFLTICLYLCLFIIRSHRLFFVNFVEYEK 143  
QY 116 IYLDGDIQVFDNIDHFLDLPNGQFYAVMDCFCEKTSWSPQYKIGYCCQCPDKVTWPEAK 175  
Db 144 IYLDGDIQVFSNIDHFLDTPRGYLYAVKDCFCEISWSKTPQPKIGYCCQCPDKVTWPEAK 203  
QY 176 LGPKPPLYFNAGMFVYEPNLSTYHNLLLETVKIVPPTLFAEQDFLNMFFKDIYKPIPPVYN 235  
Db 204 LGSPPPVYFNAGMLVFEPNLLTYEDLLRVVQITPTTYFAEQDFLNEYFTDIYKPIPSY 263  
QY 236 LVLAMLRHPENIELDQVKVVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDES 295  
Db 264 LVMAMLRHPENIDLDQISVIHYCANGSKPWRFTETEHEMDREDIKMLVKKWWDIYEDSS 323  
QY 296 LDYKNVVIGDSHKKQOTLQQFIEALSEAGALQYVKAPSAA 335  
Db 324 LDYKNFV--ETESKLSPINATLASKESVGDVLIISLAPSAA 361

RESULT 7  
T02295  
hypothetical protein T13D8.32 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 22-Oct-1999  
C;Accession: T02295  
R;Vysotskaia, V.S.; Schwartz, J.R.; Kwan, A.; Toriumi, M.; Yu, G.; Oji, O.; Liu, S.;  
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel  
submitted to the EMBL Data Library, June 1998  
A;Description: Arabidopsis thaliana chromosome 1 BAC T13D8 sequence.  
A;Reference number: Z14649  
A;Accession: T02295  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-345 <VYS>  
A;Cross-references: EMBL:AC004473; NID:g3108025; PID:g3249091; GSPDB:GN00059; ATSP:T1  
C;Genetics:  
A;Gene: ATSP:T13D8.32  
A;Map position: 1  
A;Introns: 119/3; 227/3; 272/3

Query Match 61.1%; Score 1110; DB 2; Length 345;  
Best Local Similarity 59.9%; Pred. No. 1.8e-84;  
Matches 203; Conservative 46; Mismatches 64; Indels 26; Gaps 3;  
QY 20 EKRAYVTFLACTGDYVKGVVGLAKGLRKAASKYPLVAVLPDVPEDHRKQLVDQGVVKE 79  
Db 10 KERAYVTFLAGNGDYVKGVVGLAKGLRKAASKYPLVAVLPDVPEDHRQILLAQGCIIREI 69

QY 80 IEPVYPENQTEFAMAYVYVINYSKLRW-----EFVEYNKMIYL 118  
Db 70 IEPVHPDSQDAYARAYIINYSKLRWVSVIYRLHENESLRLSLNNFEEYNKMIYL 129  
QY 119 DGDIOVFDNIDHFLDLPNGQFYAVMDCFCCEKTSWSPQYKIGYCQCPDKVTWPEAKLGP 178  
Db 130 DADIOVFGNIDDLFDMDQGYLHGVLSFCCEKIWSYTPLYSIGYCQCPDKVWPWPAEMESA 189  
QY 179 KPPLYFNAGMFVYEPNLSYHNLLTETVKIYVPTLFAEQDFLNMFKDIYKPPVYNLVL 238  
Db 190 PPSPYFNAGMFVYEPNLSYHNLLTETVKIYVPTLFAEQDFLNMFKDKVFPVYNLIL 249  
QY 239 AMLWRHPENIELDQVQVHYCAAGAKPWRFTGEENMDREDIKMLVKKWMDIYNDESLDY 298  
Db 250 SVLWRHPGKVDLESYVHYCPGSKPWRYTGEENMDREDYKMLIKKWMIDIYNDESLEDF 309  
QY 299 KNVVIGDSHKKQTLQ--FIEALSEAGALQYVKAPSAA 335  
Db 310 KPKSPAD--LEATVLESTIIASVTEAPLSYSPAAPSAA 345  
RESULT 8  
T00444  
hypotheical protein T14N5.1 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 22-Oct-1999  
C;Accession: T00444  
R;Fiederspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo,  
; Vysotskaia, V.S.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.  
submitted to the EMBL Data Library, September 1998  
A;Reference number: Z14152  
A;Accession: T00444  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1201 <FED>  
A;Cross-references: EMBL:AC004260; NID:g3176694; PID:g3540195; GSPDB:GN00059; ATSP:T14N5  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Gene: ATSP:T14N5.1  
A;Map position: 1  
A;Introns: 251/1; 569/1; 601/2; 858/1  
Query Match 14.2%; Score 258.5; DB 2; Length 1201;  
Best Local Similarity 24.2%; Pred. No. 4.5e-13;  
Matches 85; Conservative 50; Mismatches 117; Indels 99; Gaps 12;  
QY 4 EINTKLTVPVHS-----ATGGEKRAYVTFLAGTDYKGVVGLAKGLRKAK 49  
Db 836 QLRQKLQPLVPGSCELSVPLQAKDNFYSGAKKPAYATILHSAQFYVCGAIAAAQSRMSG 895  
QY 50 SKYPLVAVLDPVPEDHRKQLVDQGCVVKEIEPVYPENQTEFAMAYVYVINYSKLRWIF 109  
Db 896 STRDLVILVDEITSEYHKSGLVAAGWKIQMFORI--RNPNAVNPAYNEWNYSKFRLWQL 952  
QY 110 VEYNKMIYLDGDIQVFDNIDHFLDLPNGQFYAVMDCFCCEKTSWSPQYKIGYCQCPDKV 169  
Db 953 TEYSKIIFIDADMILRLNIDFLFE----- 976  
QY 170 TWPEAKLGPKPPLYFNAGMFVYEPNLSYHNLLTETVKIYVPTLFAEQDFLNMFKDIYKP 229  
Db 977 -FPEISATGNATLFLNSGLMVVEPSNSTFQLLMDNINEVSVYNGDQGYLNEIFTWHR- 1034  
QY 230 IPPVYNLVLAMLRHPENIELDQVK-----VVHYCAAGAKPW----- 266  
Db 1035 IPKHMNF-LKHFWEGDEP-EIKKMKTSLFGADPPILYVLHYLGYN-KPWLCFRDYDCNWN 1091  
QY 267 -----RFTGEEENMDREDIKMLVKKWMDIYN--DESLDYKNVIGDSHKKQ 310  
Db 1092 VDIQFEFASDEAH-----KTWVRVHDAMPENL-HKFCLLRSKQKAQ 1131

RESULT 9  
JC4695  
glycogenin glucosyltransferase (EC 2.4.1.186) - human  
N;Alternate names: glycogenin  
N;Contains: glycogen(starch) synthase, glycogenin subunit  
C;Species: Homo sapiens (man)  
C;Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 11-Jun-1999  
C;Accession: JC4695; G01948; S45141  
R;Barbetti, F.; Rocchi, M.; Bossolasco, M.; Cordera, R.; Sbraccia, P.; Finelli, P.; C  
Biochem. Biophys. Res. Commun. 220, 72-77, 1996  
A;Title: The human skeletal muscle glycogenin gene: cDNA, tissue expression, and chro  
A;Reference number: JC4695; MUID:96176958  
A;Accession: JC4695  
A;Molecule type: mRNA  
A;Residues: 1-333 <BAR>  
A;Cross-references: GB:U44131; NID:g1174166; PIDN:AAB00114.1; PID:g1174167  
A;Experimental source: skeletal muscle  
R;Rodriguez, I.R.  
submitted to the EMBL Data Library, July 1995  
A;Reference number: G08914  
A;Accession: G01948  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-333 <ROD>  
A;Cross-references: EMBL:U31525; NID:g976399; PIDN:AAB09752.1; PID:g976400  
R;Leffers, H.; Wiemann, S.; Ansoerge, W.  
submitted to the EMBL Data Library, June 1994  
A;Description: Cloning and sequencing of a cDNA encoding human glycogenin.  
A;Reference number: S45140  
A;Accession: S45141  
A;Molecule type: mRNA  
A;Residues: 1-203, 'KMSQEPYHICPLGRSQLWHSRLYPRKNR', 'NDGNRRLIWEQIPLTTSRGNLTSSR', 'NT  
A;Cross-references: EMBL:X79537; NID:g496894; PIDN:CAA56073.1; PID:g496895  
C;Genetics:  
A;Gene: GDB:GYG  
A;Cross-references: GDB:1326953  
A;Map position: 3q24-3q25.1  
C;Function:  
A;Description: catalyzes the alpha-glucosylation of itself on a specific tyrosine res  
osyl-glycogenin by UDPglucose to produce glycogen and UDP  
A;Pathway: glycogen/starch biosynthesis  
A;Note: required to initiate the synthesis of glycogen  
C;Superfamily: glycogenin  
C;Keywords: acetylated amino end; glycogen/starch biosynthesis; glycoprotein; glycosy  
F;2/Modified site: acetylated amino end (Thr) (in mature form) #status predicted  
F;195/Binding site: carbohydrate (Tyr) (covalent) #status predicted  
Query Match 13.3%; Score 241; DB 1; Length 333;  
Best Local Similarity 24.8%; Pred. No. 2.4e-12;  
Matches 78; Conservative 43; Mismatches 120; Indels 74; Gaps 11;  
QY 22 RAYVTELAGTDYKGVVGLAKGLRKAKSKYPLVAVLDPVPEDHRKQL--VDQGCVVKE 79  
Db 4 QAFVT-LTNDAYAKGALVGLGSLKQHRTRRLVVLATPQVSDSMRKVLETVFDEVIMVD 62  
QY 80 IEPVYPENQTEFAMAYVYVINYSKLRWIFVEYKMYLDGDIQVFDNIDHFLDLPNGQF 139  
Db 63 VLDSDGSAHLTLMKRPELGVTLTKLHCWSLTQYKCVFMDADTLVLANIDDLDFREE--- 119  
QY 140 YAVMDCFCCEKTSWSPQYKIGYCQCPDKVTWPEAKLGPKPPLYFNAGMFVYEPNLSYH 199  
Db 120 -----LSAAPDP-GWPDG-----FNSGVFYQPSVETYN 147  
QY 200 NLETVKIVPPTLFAEQDFLNMFK-----DIYKPIPPVYNLVLAMLRHPENIEL--DQ 252  
Db 148 QLLHLASEQGSFDDGGQDQILNTEFFSSWATTDIRKHLPIYNLSSISYSLPAKVFEGAS 207  
QY 253 KVYVHYCAAGAKPWRFT-----GEEENMDREDIKMLVKKWMDIYNDESIDYKNV 301  
Db 208 AKVYVHFLGR-VKPNWNTYDPKTSVKSEAHDPNMTHPFLIL---WNIFTNVL----- 258  
QY 302 VIGDSHKKQTLQOF 316





A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-372 <WOO>  
A:Cross-references: EMBL:AL021730; PIDN:CAAL6830.1;  
A:Experimental source: strain 972h-; cosmid c4C3  
C:Genetics:  
A:Gene: SPDB:SPBC4C3.08  
A:Map position: 2

Query Match	12.4%;	Score	225.5;	DB	2;	Length	372;
Best Local Similarity	25.3%;	Pred.	No. 5.4e-11;				
Matches	89;	Conservative	44;	Mismatches	128;	Indels	91; Gaps
							18;

  

QY	2	APEINTKLTVPVHSATG---	GEKRAYVTFL--AGTGD-----	--YVKGVVGLAKGL	45
		:::: ::	: : :  : :		
Db	38	SPEIQRS----	VYTTLGLAPSSKMAFVTMLTVRAANGENEVENTQQDWYYNSTRLLVHRL	93	
QY	46	RK---AKSKYPLVAVLDPDVPEDHRKQLVDQGCVVKEIEPVYPPE----	NQTEFAMAYY	97	
		:  :	: : :  : : :  :	:	
Db	94	VKFEPKTSKYPPVVLAMKGIDQWKLDQLQEDGAIVKVVDPLYAHEVVDDVNDIALLDSRW	153		
QY	98	VINYSKLRIWEFVEYNKMIYLDGGDIQVFDNIDLHFDL-----	PNGQFYAV-MD	144	
		: : : : :  : : :  : : :  :			
Db	154	SMMFTKLRVFEMYEYDRICFLDSDILPIKKMDKVFVDHQLSYKSDSVLFPPTLFYKPPRS	213		
QY	145	CFCEKTWSHSPQYKIGYCQCQP-----	-DKVTWPEAKLGPKPPL--YFNAGMFVYEPNL	195	
		:   :		: : :  :	
Db	214	IFWRRTETEEFAAYGLTRDDLYPYVFAAVSDPGMWHET----	PPPFKDIFYNAGLFFVKPLK	269	
QY	196	STYHNLLLETVKIVPPTLF-----	AEQDFLNMYFKDIYKPIPPVYNLVLAMLWRH-----	244	
		:   : :			
Db	270	AHYKRLMALARF--PKLYDNANMMEQSLLNF-----	-AYNSAGAFPWESLDWTEN	316	
QY	245	---PENIELDQVKVVHYCAAGAKPWRFTEEENMDREDIKMLVKKKWDIYND	293		
		:   : : :  :		:	
Db	317	GLWARKNDLPYLKAVH-----	GKHWOPEG-SLGYDEDTSKL-----WWDAFAQE	358	

RESULT 13  
A45729  
glycogenin homolog, sqdB 3'-region - Rhodobacter sphaeroides  
C:Species: Rhodobacter sphaeroides  
C:Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 08-Oct-1999  
C:Accession: A45729; S27663  
R:Benning, C.; Somerville, C.R.  
J. Bacteriol. 174, 6479-6487, 1992  
A:title: Identification of an operon involved in sulfolipid biosynthesis in Rhodobacter  
A:Reference number: A45729; MUID:93015699  
A:Accession: A45729  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-260 <BEN>  
A:Cross-references: EMBL:M89780; NID:g152038; PIDN:AAA73224.1; PID:g152040  
C:Genetics:  
A:Start codon: GTG

	Query Match	11.3%;	Score 205.5;	DB 2;	Length 260;
	Best Local Similarity	23.2%;	pred. No. 1.6e-09;		
	Matches 66;	Conservative 46;	Mismatches 93;	Indels 79;	Gaps 12;
QY	33	DYKGVVGLAKGLRKA	SKYPLVAVLPDVPED	HRRKQLVDQGVVKEIE--	PVYPPE
Db	5	DYALGARALLRSL	SGTTADRVV-LHTDV	PEEALPLRALGARLVR	VELLPTSP
QY	91	EFAMAYY-----	-----VINYSKLRI	WEEFVEYNKMIYLDG	DIQVFDNIDHLFDL
Db	64	HAREALHARAAFT	KGGKPPHTPLDNF	AKRLWLQVDYRSV	VIDADALVLRNVD
QY	135	PNGQFYAVMDC	CFCEKTSWHS	PQYKIGYCCQCPDK	VKTWTWPEAKLGPK
Db	124	PE-----FC-----	-----AAPN	VIYESLSDFHRMNS	144

QY 187 GMFVYEPNLSTYHNLLLETVKIVPTLF--AEQDFLNMTFKDIYPVPYNLVLA MLWR 243  
| : | | : | | : | | : | | : | | : | | : |  
Db 145 GVFTARPSTDTYARMLEALDV--PGAFWRRTDQSFLQQFPD-WQLPVCNNMLQYVWFA 201

QY 244 HPENIELDQKVVVHYCAAGAKPWRTGEEENMDREDIKMLVKKW 287  
| : | | : | | : | | : | | : | | : | | : |  
Db 202 MPELWSWEQIRILHF-QYEKPWQ---AHDKADR--LRPLIDLW 238

RESULT 14  
T05984  
hypothetical protein F17M5.90 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cross)  
C;Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
C;Accession: T05984  
R;Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.  
submitted to the Protein Sequence Database, March 1999  
A;Reference number: Z15263  
A;Accession: T05984  
A;Molecule type: DNA  
A;Residues: 1-333 <BEV>  
A;Cross-references: EMBL:AL035678; GSPDB:GN00062; ATSP:F17M5.90  
A;Experimental source: cultivar Columbia; BAC clone F17M5  
C;Genetics:  
A;Gene: ATSP:F17M5.90  
A;Map position: 4  
A;Introns: 193/2

[illegible]

RESULT 15 -  
T31001  
hypothetical protein F56B6.4 - *Caenorhabditis elegans*  
C;Species: *Caenorhabditis elegans*  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 29-Oct-1999  
C;Accession: T31001  
R;Stellyes, L.; Stellye, L.  
submitted to the EMBL Data Library, September 1999  
A;Description: The sequence of *C. elegans* cosmid F56B6.  
A;Reference number: Z20957  
A;Accession: T31001  
A;Status: preliminary; translated from GB/EMBL/DBBJ  
A;Molecule type: DNA  
A;Residues: 1-429 <STE>  
A;Cross-references: EMBL:U64599; PIDN:AAB04562.1  
A;Experimental source: strain Bristol N2  
C;Genetics:  
A;Map position: X

Search completed: July 2, 2002, 12:09:34  
Job time: 181 sec







Db 2 MGNVSSE-----KKALAAAKRRAYVTFLAGDGYWGVVGLAKGLRVRSAAYPLVAVLP 57  
QY 61 DVPEDHRKQLVDQGCVVKEIEPVPPENQTEFAMAYYVINYSKLRWFEVEYNKMIYLDG 120  
Db 58 DVPGEHRKRLVEQGCVVREIQVPYPPESQTFAMAYYVINYSKLRWFEVEYERMVYLD 117  
QY 121 DIQVFDNIDHFLDPLNGQFYAVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPEAKLGPKP 180  
Db 118 DIQVFDNIDHFLDLDKAFYAVKDCFCCKTWSHSPQYDYGCCQCPDEVAWPERELGPPP 177  
QY 181 PLYFNAGMFVYEPNLSTYHNLLTETKIVPPPTLFAEQDFLNMFKDIYKPIPPVYNLVLM 240  
Db 178 PLYFNAGMFVHEPGLGTAKDLDDALVVTPTTFAEQDFLNMFFREQYKPIPNVYNLVLM 237  
QY 241 LWRHPENIELDQVKVHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDESLDYKN 300  
Db 238 LWRHPENVLDQVKVHYCAAGSKPWRFTGKEENMNREDIKMLVKRWWDIYNDESLDYKE 297  
QY 301 VVIGDSHKKQOTLQOFIEALSEAGALQYVVKAPSA 335  
Db 298 -----EEDNADEASQPMRTALAEAGAVKYFPAPSA 328

RESULT 2  
US-09-073-297-2  
; Sequence 2, Application US/09073297  
; Patent No. 625114  
; GENERAL INFORMATION:  
; APPLICANT: LIGHTNER, JONATHAN EDWARD  
; TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: U.S.A.  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/073,297  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/852615  
; FILING DATE: MAY 7, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAJARIAN, WILLIAM R.  
; REGISTRATION NUMBER: 41,173  
; REFERENCE/DOCKET NUMBER: BB-1083-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-4926  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 346 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-073-297-2

Query Match 14.6%; Score 265.5; DB 4; Length 346;  
Best Local Similarity 27.6%; Pred. No. 3.6e-21;  
Matches 81; Conservative 45; Mismatches 98; Indels 69; Gaps 11;  
QY 21 KRAYVTFLAGTDGYKGVVGLAKGLRKAASKYPLVAVLPDVPEDHRKQLVDQGCVVKEI 80  
Db 7 REAYATILHSASEYVCGAITAAQSIROAGSTRDLVILVDDTISDHHKGLGLESAGWKVRII 66

QY 81 EPVYPPENQTEFAMAYYVINYSKLRWFEVEYNKMIYLDGDIQVFDNIDHFLDPLNGQFY 140  
Db 67 QLRNPKAERD---AYNEWNSKFLRLQLTLDYDKVIFIDADLLILRNIDFLFALP----- 118  
QY 141 AVMDCFCCKTWSHSPQYKIGYCCQCPDKVTWPEAKLGPKPPLYFNAGMFVYEPNLSTYHN 200  
Db 119 -----EIT-----ATGNNATL-----FNSGVMVLEPSNCTFRL 146  
QY 201 LLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVLMWLRHPENIELDQVK----- 254  
Db 147 LMEHIDEITSYNGGQGYLNEIFTWWHR-IPKHMNF-LKHFWEQDE-----EEVKAKKTRL 200  
QY 255 -----VVHYCAAGAKPWR-FTGEEENMDREDIKMLVK-----KWWDIYN 292  
Db 201 FGANPPVLYVLHY--LGRKPWLCFRDYDCNWNVEILREFASDVAHARWVKVHN 251

RESULT 3  
US-09-073-297-3  
; Sequence 3, Application US/09073297  
; Patent No. 625114  
; GENERAL INFORMATION:  
; APPLICANT: LIGHTNER, JONATHAN EDWARD  
; TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: U.S.A.  
; ZIP: 19898  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/073,297  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/852615  
; FILING DATE: MAY 7, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MAJARIAN, WILLIAM R.  
; REGISTRATION NUMBER: 41,173  
; REFERENCE/DOCKET NUMBER: BB-1083-B  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-992-4926  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 333 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-073-297-3

Query Match 13.3%; Score 241; DB 4; Length 333;  
Best Local Similarity 24.8%; Pred. No. 1.9e-18;  
Matches 78; Conservative 43; Mismatches 120; Indels 74; Gaps 11;  
QY 22 RAYVTFLAGTDGYKGVVGLAKGLRKAASKYPLVAVLPDVPEDHRKQL--VDQGCVVKE 79  
Db 4 QAFVT-LTTNDAYAKGALVGLSSLKQHRTRRLVVLATPQVSDSMRKVLETVEDEVIMVD 62  
QY 80 IEPVYPPENQTEFAMAYYVINYSKLRWFEVEYNKMIYLDGDIQVFDNIDHFLDPLNGQF 139  
Db 63 VLDSGDSAHLTLMKRPELGVTLTKLHCWSLTQYSKCVFMDADTLVLANIDDLDFREE--- 119

QY 140 YAVMDCFEKTSWSPQYKIGYCCQCPDKVTWPEAKLGPKPPLYFNAGMFVYEPNLSYH 199  
 Db 120 -----LSAAPDP-GWPDG-----FNSGVFVYQPSVETYN 147  
 QY 200 NLETVKIVPPTLFAEQDFLNMFK-----DIYKPIPPVYNLVLMLWRHPENIEL--DQ 252  
 Db 148 QLLHLASEQSGDFGQDQILNTFFSSWATDIRKHLPLFIYNLSSISYSYLPFAKVFEGAS 207  
 QY 253 VKVVHYCAAGAKPWRT-----GEENMDREDIKMLVKKWDIYNDESLDYKNV 301  
 Db 208 AKVVHFLGR-VKPNWNTYDPKTKSVKSEAHPNMTHTPEFLIL---WNIFTTNVL----- 258  
 QY 302 VIGDSHKKQQTQQF 316  
 Db 259 -----PLLQQF 264

RESULT 4  
 US-09-073-297-5  
 ; Sequence 5, Application US/09073297  
 ; Patent No. 6255114  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LIGHTNER, JONATHAN EDWARD  
 ; TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES  
 ; NUMBER OF SEQUENCES: 26  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
 ; STREET: 1007 MARKET STREET  
 ; CITY: WILMINGTON  
 ; STATE: DELAWARE  
 ; COUNTRY: U.S.A.  
 ; ZIP: 19898  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: DISKETTE, 3.50 INCH  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
 ; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/073,297  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/852615  
 ; FILING DATE: MAY 7, 1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MAJARIAN, WILLIAM R.  
 ; REGISTRATION NUMBER: 41,173  
 ; REFERENCE/DOCKET NUMBER: BB-1083-B  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 302-992-4926  
 ; TELEFAX: 302-773-0164  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 566 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; IMMEDIATE SOURCE:  
 ; CLONE: r10n.pk0027.f11  
 US-09-073-297-5

Query Match 12.5%; Score 227; DB 4; Length 566;  
 Best Local Similarity 22.3%; Pred. No. 1.5e-16;  
 Matches 69; Conservative 53; Mismatches 91; Indels 96; Gaps 10;  
 QY 15 SATGGEKRAYVTFLAGTGDYKGVVGLAKGLRKAASKYPLVAVLDPDVPEDHRKQLVDQ 74  
 Db 271 SSFTAQRVAVVTLHSSEYVCGAIALAQISQSGSTKDMILLHDDISITNISLGLSLAG 330  
 QY 75 CVVKEIEPVYPPEQTEFAMAYVINYSKLRIWEFVEYNKMIYLDGIQVFDNIDHLFDL 134  
 Db 331 WKLRVRIRSPFSKKR---SYNEWNYSKLRVWQVTDYDKLVFIDADFIIVKNIDYLF-- 385

QY 135 PNGOFYAVMDCFEKTSWSPQYKIGYCCQCPDKVTWPEAKLGPKPPLYFNAGMFVYEPN 194  
 Db 386 -----SYPLSAAGNNKVLNFGVMVLEPS 410  
 QY 195 LSTYHNL-LETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVLMLWRH----- 244  
 Db 411 ACLFEDLMKSKFKIGSYN-GDQGFLENYF-----VWVHRLSKRLNTM 452  
 QY 245 -----PENIELDQVKVVHYCAAGAKPW-FTGEEENMDREDIKMLVK--- 285  
 Db 453 KYFGDESRRHDKARNLPENLE-----GIHY--LGLKPWRICYRDYDCNWDLKRTRYASESV 505  
 QY 286 --KWWDIYN 292  
 Db 506 HARWVKVYD 514

Query Match 12.4%; Score 225; DB 4; Length 71;  
 Best Local Similarity 59.5%; Pred. No. 1e-17;  
 Matches 44; Conservative 13; Mismatches 13; Indels 4; Gaps 1;  
 QY 1 MAPEINTKLTVPVHSATGGEKRAYVTFLAGTGDYKGVVGLAKGLRKAASKYPLVAVL 60  
 Db 2 MGNPVSSSE---KKALAAAKRRAYVTFLAGDGDYKGVVGLAKGLRRLRRYSAYPLVAVL 57  
 QY 61 DVPEDHRKQLVDQ 74

RESULT 5  
 US-09-073-297-14  
 ; Sequence 14, Application US/09073297  
 ; Patent No. 6255114  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LIGHTNER, JONATHAN EDWARD  
 ; TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES  
 ; NUMBER OF SEQUENCES: 26  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY  
 ; STREET: 1007 MARKET STREET  
 ; CITY: WILMINGTON  
 ; STATE: DELAWARE  
 ; COUNTRY: U.S.A.  
 ; ZIP: 19898  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: DISKETTE, 3.50 INCH  
 ; COMPUTER: IBM PC COMPATIBLE  
 ; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
 ; SOFTWARE: MICROSOFT WORD VERSION 7.0A  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/073,297  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/852615  
 ; FILING DATE: MAY 7, 1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: MAJARIAN, WILLIAM R.  
 ; REGISTRATION NUMBER: 41,173  
 ; REFERENCE/DOCKET NUMBER: BB-1083-B  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 302-992-4926  
 ; TELEFAX: 302-773-0164  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 71 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: not relevant  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; IMMEDIATE SOURCE:  
 ; CLONE: r10n.pk0027.f11  
 US-09-073-297-14









Journal of Management Inquiry 22(4) 409–424 © The Author(s) 2013. Reprints and permissions: [sagepub.com/journalsPermissions.nav](http://sagepub.com/journalsPermissions.nav) DOI: 10.1177/1056492613505501

Search completed: July 2, 2002, 12:08:57  
Job time: 174 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 12:08:26 ; Search time 53.7 Seconds  
(without alignments)  
692.918 Million cell updates/sec

Title: US-09-810-506-2  
Perfect score: 1816  
Sequence: 1 MAPEINTKLTPVHSATGGE.....FIEALSEAGALQYVKAPSAA 335

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802:\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1390.5	76.6	344	21 AAG14525	Arabidopsis thaliana
2	1390.5	76.6	344	21 AAG45747	Arabidopsis thaliana
3	1389.5	76.5	333	21 AAG14526	Arabidopsis thaliana
4	1389.5	76.5	333	21 AAG45748	Arabidopsis thaliana
5	1360.5	74.9	330	14 AAR31580	Sequence encoded b
6	1357.5	74.8	328	14 AAR31581	Sequence encoded b
7	1319	72.6	328	19 AAW82560	Rice glycogenin-li
8	1319	72.6	328	22 AAE05241	Rice water stress-
9	1222.5	67.3	279	21 AAG14527	Arabidopsis thaliana
10	1222.5	67.3	279	21 AAG45749	Arabidopsis thaliana
11	496	27.3	117	22 AAB72462	Partial WSIP. Ory

12	265.5	14.6	346	19 AAW82556	Corn glycogenin cl
13	265.5	14.6	346	22 AAE05237	Corn glycogenin pr
14	248	13.7	307	22 ABB64446	Drosophila melanog
15	248	13.7	614	21 AAG39868	Arabidopsis thaliana
16	248	13.7	659	21 AAG39867	Arabidopsis thaliana
17	241	13.3	333	19 AAW82557	Human glycogenin p
18	241	13.3	333	22 AAE05238	Human glycogenin p
19	241	13.3	333	22 AAB48187	Human glycogenin p
20	240	13.2	469	19 AAW75392	Human liver glyco
21	240	13.2	501	19 AAW75391	Human liver glyco
22	235	12.9	461	19 AAW75393	Human liver glyco
23	227	12.5	566	19 AAW82559	A. thaliana homolo
24	227	12.5	566	22 AAE05240	Arabidopsis thaliana
25	225	12.4	71	19 AAW82563	Rice glycogenin cl
26	225	12.4	71	22 AAE05244	Rice water stress
27	214.5	11.8	353	22 AAU33096	Novel human secret
28	212.5	11.7	342	19 AAW75394	Human brain glyco
29	211.5	11.6	298	19 AAW75398	Human breast glyco
30	208.5	11.5	279	21 AAY54082	Enzyme EPS8 involv
31	208.5	11.5	279	21 AAY43784	Amino acid sequenc
32	194	10.7	300	19 AAW82558	C. elegans glyco
33	194	10.7	300	22 AAE05239	Caenorhabditis ele
34	187	10.3	43	19 AAW82566	Wheat glycogenin c
35	187	10.3	43	22 AAE05247	Wheat glycogenin p
36	167	9.2	157	21 AAB28193	Human glycogenin.
37	161	8.9	132	19 AAW82564	Rice glycogenin cl
38	161	8.9	132	22 AAE05245	Rice glycogenin pr
39	145.5	8.0	93	19 AAW82562	Corn glycogenin cl
40	145.5	8.0	93	22 AAE05243	Corn glycogenin pr
41	140	7.7	454	21 AAG31966	Arabidopsis thaliana
42	140	7.7	470	21 AAG31965	Arabidopsis thaliana
43	140	7.7	506	21 AAG40552	Arabidopsis thaliana
44	140	7.7	522	21 AAG40551	Arabidopsis thaliana
45	134	7.4	306	17 AAR91313	N. gonorrhoeae gly

ALIGNMENTS

RESULT 1  
AAG14525  
ID AAG14525 standard; Protein; 344 AA.  
XX  
AC AAG14525;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 14420.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.







DE	Arabidopsis thaliana protein fragment SEQ ID NO: 14421.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;	PR	22-JUN-1999;
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	PR	23-JUN-1999;
KW	termination sequence.	PR	24-JUN-1999;
XX		PR	28-JUN-1999;
OS	Arabidopsis thaliana.	PR	29-JUN-1999;
XX		PR	30-JUN-1999;
PN	EP1033405-A2.	PR	01-JUL-1999;
XX		PR	01-JUL-1999;
PD		PR	02-JUL-1999;
XX		PR	06-JUL-1999;
PF	06-SEP-2000.	PR	06-JUL-1999;
XX		PR	08-JUL-1999;
PR	25-FEB-2000; 2000EP-0301439.	PR	09-JUL-1999;
PR		PR	12-JUL-1999;
PR	25-FEB-1999;	PR	13-JUL-1999;
PR	05-MAR-1999;	PR	14-JUL-1999;
PR	09-MAR-1999;	PR	15-JUL-1999;
PR	23-MAR-1999;	PR	16-JUL-1999;
PR	25-MAR-1999;	PR	16-JUL-1999;
PR	29-MAR-1999;	PR	19-JUL-1999;
PR	01-APR-1999;	PR	19-JUL-1999;
PR	06-APR-1999;	PR	19-JUL-1999;
PR	08-APR-1999;	PR	19-JUL-1999;
PR	16-APR-1999;	PR	19-JUL-1999;
PR	19-APR-1999;	PR	20-JUL-1999;
PR	21-APR-1999;	PR	20-JUL-1999;
PR	23-APR-1999;	PR	20-JUL-1999;
PR	28-APR-1999;	PR	21-JUL-1999;
PR	30-APR-1999;	PR	21-JUL-1999;
PR	04-MAY-1999;	PR	22-JUL-1999;
PR	05-MAY-1999;	PR	22-JUL-1999;
PR	06-MAY-1999;	PR	22-JUL-1999;
PR	06-MAY-1999;	PR	22-JUL-1999;
PR	07-MAY-1999;	PR	23-JUL-1999;
PR	11-MAY-1999;	PR	23-JUL-1999;
PR	14-MAY-1999;	PR	23-JUL-1999;
PR	14-MAY-1999;	PR	26-JUL-1999;
PR	14-MAY-1999;	PR	27-JUL-1999;
PR	14-MAY-1999;	PR	27-JUL-1999;
PR	18-MAY-1999;	PR	27-JUL-1999;
PR	19-MAY-1999;	PR	28-JUL-1999;
PR	20-MAY-1999;	PR	02-AUG-1999;
PR	21-MAY-1999;	PR	02-AUG-1999;
PR	24-MAY-1999;	PR	02-AUG-1999;
PR	25-MAY-1999;	PR	03-AUG-1999;
PR	27-MAY-1999;	PR	04-AUG-1999;
PR	28-MAY-1999;	PR	04-AUG-1999;
PR	01-JUN-1999;	PR	05-AUG-1999;
PR	03-JUN-1999;	PR	05-AUG-1999;
PR	04-JUN-1999;	PR	06-AUG-1999;
PR	07-JUN-1999;	PR	06-AUG-1999;
PR	08-JUN-1999;	PR	09-AUG-1999;
PR	10-JUN-1999;	PR	09-AUG-1999;
PR	10-JUN-1999;	PR	10-AUG-1999;
PR	14-JUN-1999;	PR	11-AUG-1999;
PR	16-JUN-1999;	PR	12-AUG-1999;
PR	16-JUN-1999;	PR	13-AUG-1999;
PR	17-JUN-1999;	PR	13-AUG-1999;
PR	18-JUN-1999;	PR	16-AUG-1999;
PR	18-JUN-1999;	PR	17-AUG-1999;
PR	18-JUN-1999;	PR	18-AUG-1999;
PR	18-JUN-1999;	PR	20-AUG-1999;
PR	18-JUN-1999;	PR	20-AUG-1999;
PR	18-JUN-1999;	PR	20-AUG-1999;
PR	18-JUN-1999;	PR	23-AUG-1999;
PR	18-JUN-1999;	PR	23-AUG-1999;
PR	18-JUN-1999;	PR	25-AUG-1999;
PR	18-JUN-1999;	PR	26-AUG-1999;
PR	18-JUN-1999;	PR	27-AUG-1999;
PR	21-JUN-1999;	PR	27-AUG-1999;

99US-0139899.	
99US-0140353.	
99US-0140354.	
99US-0140695.	
99US-0140823.	
99US-0140991.	
99US-0141287.	
99US-0141842.	
99US-0142154.	
99US-0142055.	
99US-0142390.	
99US-0142803.	
99US-0142920.	
99US-0142977.	
99US-0143542.	
99US-0143624.	
99US-0144005.	
99US-0144085.	
99US-0144086.	
99US-0144325.	
99US-0144331.	
99US-0144332.	
99US-0144333.	
99US-0144334.	
99US-0144335.	
99US-0144352.	
99US-0144632.	
99US-0144884.	
99US-0144814.	
99US-0145086.	
99US-0145088.	
99US-0145085.	
99US-0145087.	
99US-0145089.	
99US-0145192.	
99US-0145145.	
99US-0145218.	
99US-0145224.	
99US-0145276.	
99US-0145913.	
99US-0145918.	
99US-0145919.	
99US-0145951.	
99US-0146386.	
99US-0146388.	
99US-0146389.	
99US-0147038.	
99US-0147204.	
99US-0147302.	
99US-0147192.	
99US-0147260.	
99US-0147303.	
99US-0147416.	
99US-0147493.	
99US-0147935.	
99US-0148171.	
99US-0148319.	
99US-0148341.	
99US-0148565.	
99US-0148684.	
99US-0149368.	
99US-0149175.	
99US-0149426.	
99US-0149722.	
99US-0149723.	
99US-0149929.	
99US-0149902.	
99US-0149930.	
99US-0150566.	
99US-0150884.	
99US-0151065.	
99US-0151066.	
99US-0151080.	





QY 129 DHIQDLPNGQFYAVMDCFCCEKTSWSPQYKIGYCCQCPDKVTWPEAKLGPKPPLYFNAGM 188  
Db 126 dhlfdlpdgylyavmdcfccektwshtpqykyrcqcpdkvqpkaelgeppalyfnagm 185  
QY 189 FVEPNLSTYHNLLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVLAMLRHPENI 248  
Db 186 flyepnletyedllrltkitpftpfaeqdfnmfkykpiplvynlamlwrhpenv 245  
QY 249 ELQDVKVWHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDESIDYKN-VVIGDSH 307  
Db 246 elgkvkvwhycaagskpwyrtgkeanmeredikmlvkkwddiyddesldydkkptvvdte 305  
QY 308 KQOQTLQOQFIEALSEAGALQYVKAPSAA 335  
Db 306 vdlvnlkpfitalteagrlnyvtapsaa 333  
RESULT 5  
AAR31580  
ID AAR31580 standard; Protein; 330 AA.  
XX  
AC AAR31580;  
XX  
DT 06-JUN-1993 (first entry)  
XX  
DE Sequence encoded by zucchini leaf galactinol synthase cDNA.  
XX  
KW Galactinol synthase; enzyme; raffinose saccharide; sucrose.  
XX  
OS Cucurbita pepo Burpee's hybrid.  
XX  
PN W09302196-A.  
XX  
PD 04-FEB-1993.  
XX  
PF 24-JUL-1992; 92WO-US06057.  
XX  
PR 24-JUL-1991; 91US-0735066.  
XX  
PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX  
PI Becker-Manley MF, Kerr PS, Pearlstein RW, Pierce JW;  
PI Schweiger BJ;  
XX  
DR WPI; 1993-058793/07.  
DR N-PSDB; AAQ36503.  
XX  
PT Nucleotide sequence of galactinol synthase from zucchini and  
PT soybean - used to produce plants having altered levels of  
PT raffinose saccharide(s) and/or sucrose  
XX  
PS Disclosure; Pages 59-61; 80pp; English.  
XX  
CC The nucleotide 1 is the first nucleotide following the Eco RI  
CC restriction site used in cloning the cDNA, reading from 5' to 3'  
CC on the cDNA insert, and nucleotide 1265 is the last nucleotide  
CC of the cDNA insert. The inventors claim a method of varying the  
CC level of D-galactose containing oligosaccharides of sucrose in  
CC plants in response to end-user requirements which involves combining  
CC the nucleic acid fragment of AAQ36503 and AAQ36504 with suitable  
CC regulatory sequences for expression and localization in plant tissues.  
XX  
SQ Sequence 330 AA;

Query Match 74.9%; Score 1360.5; DB 14; Length 330;  
Best Local Similarity 75.6%; Pred. No. 1.3e-137;  
Matches 242; Conservative 36; Mismatches 41; Indels 1; Gaps 1;

QY 16 ATGGEKRAYVTFLAGTGDYVKGVLAKGLRRAKSKYPLVAVLPDVPEDHRKQLVDQGC 75  
Db 12 stdapkravtflagngdywkgvvgglakgrkvktvplivavlpdvpedhrqileyygc 71

QY 76 VVKELEPVYPPENQTEFAMAYVYVINYSKLRIFEVEYNKMIYLDGDIQVFNIDHLDLP 135  
Db 72 ivrelepvyppanqtqfamayvynysklriwefveyekliylgdqivfenidhlfemp 131  
QY 136 NGQFYAVMDCFCCEKTSWSPQYKIGYCCQCPDKVTWPEAKLGPKPPLYFNAGFVYEPNL 195  
Db 132 ngyfayvmdcfcektwsnspqykigycqcpdkvkpveemgnpplyfnagfvyepdl 191  
QY 196 STYHNLLETVKIVPPTLFAEQDFLNMFKDIYKPIPPVYNLVLAMLRHPENIELDQVKV 255  
Db 192 ftykdllctckattptlfaeqdfnmfndiykpiipynlvmamlwrhpenidvdkvk 251  
QY 256 VHYCAAGAKPWRFTGEEENMDREDIKMLVKKWWDIYNDESIDYKNVIGDSHKQQTLOQ 315  
Db 252 vhycaagskpwyrtgeenmdredikmlvkkwweyedesldyqnvikset-keatnvp 310  
QY 316 FIEALSEAGALQYVKAPSAA 335  
Db 311 lvsvlseaeavnhitapsaa 330  
RESULT 6  
AAR31581  
ID AAR31581 standard; Protein; 328 AA.  
XX  
AC AAR31581;  
XX  
DT 06-JUN-1993 (first entry)  
XX  
DE Sequence encoded by soybean seed galactinol synthase cDNA.  
XX  
KW Galactinol synthase; enzyme; raffinose saccharide; sucrose.  
XX  
OS Glycine max Cultivar Wye.  
XX  
PN W09302196-A.  
XX  
PD 04-FEB-1993.  
XX  
PF 24-JUL-1992; 92WO-US06057.  
XX  
PR 24-JUL-1991; 91US-0735066.  
XX  
PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX  
PI Becker-Manley MF, Kerr PS, Pearlstein RW, Pierce JW;  
PI Schweiger BJ;  
XX  
DR WPI; 1993-058793/07.  
DR N-PSDB; AAQ36504.  
XX  
PT Nucleotide sequence of galactinol synthase from zucchini and  
PT soybean - used to produce plants having altered levels of  
PT raffinose saccharide(s) and/or sucrose  
XX  
PS Disclosure; Pages 61-63; 80pp; English.  
XX  
CC The nucleotide 1 is the first nucleotide following the Eco RI  
CC restriction site used in cloning the cDNA, reading from 5' to 3'  
CC on the cDNA insert, and nucleotide 1265 is the last nucleotide  
CC of the cDNA insert. The inventors claim a method of varying the  
CC level of D-galactose containing oligosaccharides of sucrose in  
CC plants in response to end-user requirements which involves combining  
CC the nucleic acid fragment of AAQ36503 and AAQ36504 with suitable  
CC regulatory sequences for expression and localization in plant tissues.  
XX  
SQ Sequence 328 AA;

Query Match 74.8%; Score 1357.5; DB 14; Length 328;  
Best Local Similarity 74.3%; Pred. No. 2.7e-137;  
Matches 251; Conservative 29; Mismatches 45; Indels 13; Gaps 4;









PR 25-FEB-1999;	99US-0121825.	PR 13-JUL-1999;	99US-0143542.
PR 05-MAR-1999;	99US-0123180.	PR 14-JUL-1999;	99US-0143624.
PR 09-MAR-1999;	99US-0123548.	PR 15-JUL-1999;	99US-0144005.
PR 23-MAR-1999;	99US-0125788.	PR 16-JUL-1999;	99US-0144085.
PR 25-MAR-1999;	99US-0126264.	PR 16-JUL-1999;	99US-0144086.
PR 29-MAR-1999;	99US-0126785.	PR 19-JUL-1999;	99US-0144325.
PR 01-APR-1999;	99US-0127462.	PR 19-JUL-1999;	99US-0144331.
PR 06-APR-1999;	99US-0128234.	PR 19-JUL-1999;	99US-0144332.
PR 08-APR-1999;	99US-0128714.	PR 19-JUL-1999;	99US-0144333.
PR 16-APR-1999;	99US-0129845.	PR 19-JUL-1999;	99US-0144334.
PR 19-APR-1999;	99US-0130077.	PR 19-JUL-1999;	99US-0144335.
PR 21-APR-1999;	99US-0130449.	PR 20-JUL-1999;	99US-0144352.
PR 23-APR-1999;	99US-0130510.	PR 20-JUL-1999;	99US-0144632.
PR 23-APR-1999;	99US-0130891.	PR 20-JUL-1999;	99US-0144884.
PR 28-APR-1999;	99US-0131449.	PR 21-JUL-1999;	99US-0144814.
PR 30-APR-1999;	99US-0132048.	PR 21-JUL-1999;	99US-0145086.
PR 04-MAY-1999;	99US-0132407.	PR 21-JUL-1999;	99US-0145088.
PR 05-MAY-1999;	99US-0132484.	PR 22-JUL-1999;	99US-0145085.
PR 06-MAY-1999;	99US-0132485.	PR 22-JUL-1999;	99US-0145087.
PR 06-MAY-1999;	99US-0132486.	PR 22-JUL-1999;	99US-0145089.
PR 07-MAY-1999;	99US-0132487.	PR 22-JUL-1999;	99US-0145192.
PR 11-MAY-1999;	99US-0132863.	PR 23-JUL-1999;	99US-0145145.
PR 14-MAY-1999;	99US-0134256.	PR 23-JUL-1999;	99US-0145218.
PR 14-MAY-1999;	99US-0134218.	PR 23-JUL-1999;	99US-0145224.
PR 14-MAY-1999;	99US-0134219.	PR 26-JUL-1999;	99US-0145276.
PR 14-MAY-1999;	99US-0134221.	PR 27-JUL-1999;	99US-0145913.
PR 14-MAY-1999;	99US-0134370.	PR 27-JUL-1999;	99US-0145918.
PR 18-MAY-1999;	99US-0134768.	PR 27-JUL-1999;	99US-0145919.
PR 19-MAY-1999;	99US-0134941.	PR 28-JUL-1999;	99US-0145951.
PR 20-MAY-1999;	99US-0135124.	PR 02-AUG-1999;	99US-0146386.
PR 21-MAY-1999;	99US-0135353.	PR 02-AUG-1999;	99US-0146388.
PR 24-MAY-1999;	99US-0135629.	PR 02-AUG-1999;	99US-0146389.
PR 25-MAY-1999;	99US-0136021.	PR 03-AUG-1999;	99US-0147038.
PR 27-MAY-1999;	99US-0136392.	PR 04-AUG-1999;	99US-0147204.
PR 28-MAY-1999;	99US-0136782.	PR 04-AUG-1999;	99US-0147302.
PR 01-JUN-1999;	99US-0137222.	PR 05-AUG-1999;	99US-0147192.
PR 03-JUN-1999;	99US-0137528.	PR 05-AUG-1999;	99US-0147260.
PR 04-JUN-1999;	99US-0137502.	PR 06-AUG-1999;	99US-0147303.
PR 07-JUN-1999;	99US-0137724.	PR 06-AUG-1999;	99US-0147416.
PR 08-JUN-1999;	99US-0138094.	PR 09-AUG-1999;	99US-0147493.
PR 10-JUN-1999;	99US-0138540.	PR 09-AUG-1999;	99US-0147935.
PR 10-JUN-1999;	99US-0138847.	PR 10-AUG-1999;	99US-0148171.
PR 14-JUN-1999;	99US-0139119.	PR 11-AUG-1999;	99US-0148319.
PR 16-JUN-1999;	99US-0139452.	PR 12-AUG-1999;	99US-0148341.
PR 16-JUN-1999;	99US-0139453.	PR 13-AUG-1999;	99US-0148565.
PR 17-JUN-1999;	99US-0139492.	PR 13-AUG-1999;	99US-0148684.
PR 18-JUN-1999;	99US-0139454.	PR 16-AUG-1999;	99US-0149368.
PR 18-JUN-1999;	99US-0139455.	PR 17-AUG-1999;	99US-0149175.
PR 18-JUN-1999;	99US-0139456.	PR 18-AUG-1999;	99US-0149426.
PR 18-JUN-1999;	99US-0139457.	PR 20-AUG-1999;	99US-0149722.
PR 18-JUN-1999;	99US-0139458.	PR 20-AUG-1999;	99US-0149723.
PR 18-JUN-1999;	99US-0139459.	PR 20-AUG-1999;	99US-0149929.
PR 18-JUN-1999;	99US-0139460.	PR 23-AUG-1999;	99US-0149902.
PR 18-JUN-1999;	99US-0139461.	PR 23-AUG-1999;	99US-0149930.
PR 18-JUN-1999;	99US-0139462.	PR 25-AUG-1999;	99US-0150566.
PR 18-JUN-1999;	99US-0139463.	PR 26-AUG-1999;	99US-0150884.
PR 18-JUN-1999;	99US-0139750.	PR 27-AUG-1999;	99US-0151065.
PR 18-JUN-1999;	99US-0139763.	PR 27-AUG-1999;	99US-0151066.
PR 21-JUN-1999;	99US-0139817.	PR 27-AUG-1999;	99US-0151080.
PR 22-JUN-1999;	99US-0139899.	PR 30-AUG-1999;	99US-0151303.
PR 23-JUN-1999;	99US-0140353.	PR 31-AUG-1999;	99US-0151438.
PR 23-JUN-1999;	99US-0140354.	PR 01-SEP-1999;	99US-0151930.
PR 24-JUN-1999;	99US-0140695.	PR 07-SEP-1999;	99US-0152363.
PR 28-JUN-1999;	99US-0140823.	PR 10-SEP-1999;	99US-0153070.
PR 29-JUN-1999;	99US-0140991.	PR 13-SEP-1999;	99US-0153758.
PR 30-JUN-1999;	99US-0141287.	PR 15-SEP-1999;	99US-0154018.
PR 01-JUL-1999;	99US-0141842.	PR 16-SEP-1999;	99US-0154039.
PR 01-JUL-1999;	99US-0142154.	PR 20-SEP-1999;	99US-0154779.
PR 02-JUL-1999;	99US-0142055.	PR 22-SEP-1999;	99US-0155139.
PR 06-JUL-1999;	99US-0142390.	PR 23-SEP-1999;	99US-0155486.
PR 08-JUL-1999;	99US-0142803.	PR 24-SEP-1999;	99US-0155659.
PR 09-JUL-1999;	99US-0142920.	PR 28-SEP-1999;	99US-0156458.
PR 12-JUL-1999;	99US-0142977.	PR 29-SEP-1999;	99US-0156596.



PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 67.3%; Score 1222.5; DB 21; Length 279;  
Best Local Similarity 78.5%; Pred. No. 7.3e-123;  
Matches 219; Conservative 28; Mismatches 31; Indels 1; Gaps 1;

QY 58 VLPDVPEDHRKQLVDQGCVVKEIEPVYPPENQTEFAMAYVINYSLRIWEFVEYNKMIY 117  
Db 1 mlpdrvpehrrilvdqgcivreiepvypnenqtqfamayvinyksklriwkvfveyskmiy 60  
QY 118 LDGDIQVFDNIDHFLDLPNGQFYAVMDCFCEKTWSHSPQYKIGYCQCQCPDKVTWPEAKLG 177  
Db 61 ldgdiqvyenidhldfldpdylyavmcdcfeektwshtpqqykiyccqcpdkvqwpkaelg 120  
QY 178 PKPPLYFNAGMEVVEPNLSTYHNLETVKIVPPTLFAEQDFLNMVFKDIYKPIPPVYNLV 237  
Db 121 eppalyfnagmflyepnletyedllrtlkitpptpfaeqdflnmfykkiykpilpvnlv 180  
QY 238 LAMLRHPENIELDQVKVHYCAAGAKPWRFTGEENMDREDIKMLVKKWWDIYNDESID 297  
Db 181 lamlrhpnenvelgkvkvvhyaagskpwrytgkeanmeredikmlvkkwddiyydesid 240  
QY 298 YKN-VVIGDSHKKQQTLOQFIEALSEAGALQYVVKAPSAA 335  
Db 241 ykkpvtvtdtevdlnlknkpfitalteagrlnyvtapsaa 279

RESULT 11  
AAB72462  
ID AAB72462 standard; Protein; 117 AA.  
XX  
AC AAB72462;  
XX  
DT 08-MAY-2001 (first entry)  
XX  
DE Partial WSIP.  
XX UGGT; UDP-glucose:glycoprotein glucosyltransferase; enzyme; WSIP; rice.  
XX  
OS Oryza sativa.

XX WO200112845-A1.  
PN  
XX  
PD 22-FEB-2001.  
XX  
PF 27-JUL-2000; 2000WO-CA00883.  
XX  
PR 18-AUG-1999; 99US-0376330.  
XX  
PA (CANA ) NAT RES COUNCIL CANADA.  
XX  
PI Tessier DC, Dignard D, Bergeron JJM, Thomas DY;  
XX WPI; 2001-218358/22.  
DR  
XX  
PS Disclosure; Fig 10; 95pp; English.  
XX  
CC The present invention relates to a method for determining the effect of a  
test sample on UDP-glucose:glycoprotein glucosyltransferase (UGGT)  
activity. The method comprises exposing an acceptor substrate for UGGT to  
a labelled donor in the presence of the test sample and UGGT. The method  
is useful for determining UGGT activity. In particular, the method is  
useful in glucosyltransferase assay and kinetics measurement for  
determining UGGT activity. UGGT is a soluble enzyme of the endoplasmic  
reticulum which catalyses the addition of a glucose residue onto  
asparagine-linked oligosaccharides, which are present in a sequence  
folded glycoproteins. The present sequence was used in a sequence  
homology comparison with rat UGGT (see AAF60732 and AAB72436) which was  
used in the method of the present invention.  
XX  
SQ Sequence 117 AA;

Query Match 27.3%; Score 496; DB 22; Length 117;  
Best Local Similarity 72.6%; Pred. No. 3.2e-45;  
Matches 85; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 109 FVEYNKMIYLDGDIQVFDNIDHFLDLPNGQFYAVMDCFCEKTWSHSPQYKIGYCQCQCPDK 168  
Db 1 fveyermvyladadiqvfdnidhldldkgafoyavkdcfcktwshtpdydigycqqrpe 60  
QY 169 VTWPEAKLGPKPPLYFNAGMFVVEPNLSTYHNLETVKIVPPTLFAEQDFLNMVFKD 225  
Db 61 vawperelgpppplyfnagmfvhepglgtakdlldalvvtptpfaeqdflnmfire 117

RESULT 12  
AAW82556  
ID AAW82556 standard; Protein; 346 AA.  
XX  
AC AAW82556;  
XX  
DT 09-FEB-1999 (first entry)  
XX  
DE Corn glycogenin clone cc3.mn0001.f7 protein.  
XX  
KW Glycogenin; water stress protein; starch biosynthesis; corn; rice;  
wheat; plant; granules; branch chain; endosperm.  
XX  
OS Zea mays.  
XX  
FH Key Location/Qualifiers  
FT Protein 1..346  
FT /note= "Partial sequence, no start codon given"  
XX  
PN WO9850553-A1.  
XX  
PD 12-NOV-1998.











